Agricultural and Forest Meteorology 230–231 (2016) 97–113

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

journal homepage: www.elsevier.com/locate/agrformet

Upscaling evapotranspiration measurements from multi-site to the satellite pixel scale over heterogeneous land surfaces

Shaomin Liu[a,*], Ziwei Xu[a], Lisheng Song[a], Qianyi Zhao[a], Yong Ge[b], Tongren Xu[a], Yanfei Ma[a], Zhongli Zhu[a], Zhenzhen Jia[a], Fen Zhang[a]

[a] State Key Laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, Beijing 100875, China
[b] State Key Laboratory of Resources and Environmental Information System, Institute of Geographic Sciences & Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China

ARTICLE INFO

Article history:
Received 18 June 2015
Received in revised form 2 February 2016
Accepted 11 April 2016
Available online 23 April 2016

Keywords:
Evapotranspiration
Upscaling
The satellite pixel scale
The Priestley-Taylor equation

ABSTRACT

The acquisition of "ground-truth" land surface evapotranspiration (ET) data at the satellite pixel scale over heterogeneous land surfaces is crucial to develop ET estimation models and improve the accuracy of remotely sensed ET values. However, few studies have focused on methods of acquiring ET data at the satellite pixel scale. Based on multi-site eddy covariance (EC) system measurements from the "Multi-Scale Observation Experiment on Evapotranspiration over heterogeneous land surfaces" in the middle reaches of the Heihe River Basin, five upscaling methods were compared and a combined method was developed to acquire "ground-truth" ET data at the satellite pixel scale. First, this study evaluated the performances of three simple upscaling methods (the arithmetic average, area-weighted and footprint-weighted methods). The results showed that the three simple upscaling methods perform well in the relatively homogeneous pixels. For the area-weighted method, the mean absolute percentage error (MAPE) for these pixels was 6.1%. However, the accuracy was worse in the relatively heterogeneous pixels, with a MAPE of 10.8% due to the surface heterogeneity significantly affecting the accuracy of the upscaled results. Second, the upscaling of ET results from heterogeneous land surfaces at the satellite pixel scale can be significantly improved by using two upscaling methods introducing auxiliary variables (the integrated Priestley-Taylor equation method and the area-to-area regression kriging method), that can characterize the heterogeneity of the surface water and heat conditions. Finally, a combined method (applied the area-weighted method for relative homogeneous surfaces, otherwise used the method introducing auxiliary variables) was proposed to acquire both instantaneous and daily "ground-truth" ET data at the satellite pixel scale at the time of a MODIS overpass. The uncertainties of the "ground-truth" ET data were evaluated, taking the large aperture scintillometer (LAS) measurements as the satellite pixel reference. The results show that the proposed upscaling method is reasonable and feasible, and therefore could bridge the gap between in situ ET measurements and remote-sensing estimates of ET.

1. Introduction

Land surface evapotranspiration (ET) is not only a key component of surface water circulation but also an indispensable component of surface energy balances, and plays an important role in aspects of meteorology, hydrology, and ecology, etc. The high-precision acquisition of ET at the regional-to-global scale provides important scientific information that is valuable to high-interest research fields, such as global environment change, the management of basin water resources, and the sustainable development of agriculture.

With the development of new technologies since the 1970s, methods for estimating ET via remote sensing have become an effective way to obtain spatially and temporally continuous ET data at the regional scale. Consequently, these methods have attracted increasing attention from researchers and management departments and have been widely applied to various fields of study (Kalma et al., 2008; Allen et al., 2011; Wang and Dickinson, 2012). However, several uncertainties still remain in remote-sensing estimates of ET and are primarily related to model mechanisms, model inputs, parameterization schemes, and scaling issues (Jia et al.,

* Corresponding author at: State key laboratory of Remote Sensing Science, School of Geography, Beijing Normal University, No.19, Xinjiekouwai Street Beijing 100875, China.
E-mail address: smliu@bnu.edu.cn (S. Liu).

http://dx.doi.org/10.1016/j.agrformet.2016.04.008
0168-1923/

2012). The mean absolute percentage errors (MAPEs) of the current remote-sensing estimates for ET at the instantaneous, daily, monthly, and annual scales are approximately 15–30%, 14–25%, 9–35%, and 5–21%, respectively (Kalma et al., 2008; Wang and Liang, 2008; Yao et al., 2013; Velpuri et al., 2013). Therefore, remote-sensing products must be validated and calibrated based on ground measurements to provide higher-precision ET products for applications.

It is difficult to match in situ measurements with a satellite-based ET estimation using the current instrumentation (e.g., the eddy covariance (EC) system or large-aperture scintillometer (LAS)) due to the variations in the orientation, shape and size of the source areas of in situ measurements. To solve this problem, Jia et al. (2012) proposed a pixel selection method based on footprints, which was superior to the method of directly comparing in situ measurements with satellite-based estimates for a single pixel or the method of multi-pixel averaging. However, because of the spatial heterogeneity at the sub-pixel scale, the in situ measurements over heterogeneous land surfaces only approximate the "ground-truth" ET data at the satellite pixel scale (1–10 km for medium-low spatial resolutions) and can introduce errors in the validation results. If the "ground-truth" ET at the satellite pixel scale can be obtained, these problems, including spatial-scale mismatch and spatial heterogeneity at the sub-pixel scale, can be reasonably solved.

Methods for the acquisition of "ground-truth" ET data at the model grid scale (10–100 km) have been widely studied and can provide useful references for obtaining "ground-truth" ET data at the satellite pixel scale. A number of observational experiments on multi-site surface fluxes and the related parameters have been conducted over various underlying surfaces, including The First ISLSCP Field Experiment (FIFE; Sellers et al., 1988), the Hydrologic Atmospheric Pilot Experiment–Modélisation du Bilan Hydrique (HAPEX-MOBILHY; André et al., 1986), the Hydrologic Atmospheric Pilot Experiment in the Sahel (HAPEX-Sahel; Goutorbe et al., 1994), the Boardman Experiment (The Boardman Regional Flux Experiment; Doran et al., 1992), the Northern hemisphere climate Processes land surface Experiment (NOPEX; Halldin et al., 1999), Lindenberg Inhomogeneous Terrain-Fluxes between Atmosphere and Surface: a Long-term Study (LITFASS-98; Beyrich et al., 2002), and LITFASS-2003 (Beyrich and Mengelkamp, 2006). In these experiments, surface fluxes were observed at multiple sites, and surface flux values at the model grid scale were obtained using various upscaling methods. The upscaled flux values were also compared to those obtained from airborne EC systems, LASs, remote sensing estimation methods, and model simulations. The methods commonly used in previous studies of flux upscaling include the arithmetic average method (Shuttleworth et al., 1989; Peng et al., 2008), the area-weighted method (Gottschalk et al., 1999; Beyrich et al., 2006; Ezzahar et al., 2009), the footprint-weighted method (Ezzahar et al., 2007; Timmermans et al., 2009; Lu et al., 2010), numerical models (André et al., 1990; Gryning et al., 2002; Heinemann and Kerschgens, 2005), and geostatistical methods (e.g., the area-to-area kriging (ATAK) method). For example, Ge et al. (2015) proposed the area-to-area regression kriging (ATARK) method to upscale the sensible heat flux from multi-site to the satellite pixel scale over heterogeneous land surfaces and compared the results with LAS measurements.

Although the above-mentioned upscaling methods have been applied to the data analysis of various large-scale experiments conducted around the world, the upscaled results of previous studies have primarily focused on the model grid scale and have their own limitations, including how to select an optimized upscaling method according to the surface heterogeneity, and how to assess the uncertainties of the upscaled results, etc. (Shuttleworth, 1991; Heinemann and Kerschgens, 2005; Li et al., 2009).

To date, few studies have focused on methods of acquiring ET data at the satellite pixel scale, especially for medium-low spatial resolutions (1–10 km). The following issues were studied based on the multi-site measurements of the "Multi-Scale Observation Experiment on Evapotranspiration over heterogeneous land surfaces 2012 (HiWATER-MUSOEXE)" in the Zhangye area, located in the middle reaches of the Heihe River Basin: (1) three simple upscaling methods (the arithmetic average method, area-weighted method, and footprint-weighted method) to obtain ET at the satellite pixel scale were evaluated, taking LAS measurements as the satellite pixel reference; (2) for the heterogeneous condition, two upscaling methods introducing auxiliary variables (the integrated Priestley-Taylor equation method and the area-to-area regression kriging method), which can characterize surface heterogeneity, were developed and validated; and (3) both instantaneous and daily "ground-truth" ET data were acquired at the satellite pixel scale at the time of a MODIS overpass, and their uncertainties were evaluated.

2. Materials

2.1. Experiment

The HiWATER-MUSOEXE was conducted in the middle reaches of the Heihe River Basin between May and September 2012 and involved a flux observation matrix that was composed of two nested matrices: one large experimental area (30 km × 30 km) and one kernel experimental area (5.5 km × 5.5 km) (Li et al., 2013; Xu et al., 2013). In this study, the dataset observed in the 5.5 km × 5.5 km kernel experimental area of the HiWATER-MUSOEXE was used to study ET upscaling methods (Fig. 1). The kernel experimental area was located in the Yingke and Daman irrigation district, and the main surfaces were maize, residential area, vegetable, and orchard. These areas were separated into rows and columns by shelterbelts. Together, they represent the land cover and planting structure in the oasis of the middle reaches of the Heihe River Basin. Moreover, 17 elementary sampling plots were divided according to the distribution of crops, shelterbelts, residential areas, roads, and canals, as well as according to soil moisture and irrigation status. These divisions resulted in one residential area site (site 4), one orchard site (site 17), one vegetable site (site 1), and 14 maize sites. In each plot, one EC and automatic weather station (AWS) was installed to observe sensible heat flux (H), latent heat flux (LE), and meteorological elements. Two EC sets and seven layers of meteorological gradient observation systems were installed at site 15. Moreover, the transpiration of shelterbelts with different heights and diameters at breast height (DBHs) was measured using a thermal dissipation probe (TDP) around sites 6, 8, and 17. Three poplar trees in each site were each instrumented with three TDP probes installed at a height of 1.3 m. Additionally, four groups of optical large-aperture scintillometers (LASs) (eight sets with two sets in each group) were installed in the 3 × 3 and 2 × 1 MODIS pixels within the kernel experimental area (three groups in three 3 × 1 MODIS pixels, named LAS1 to LAS3 from west to east, and one group in one 2 × 1 MODIS pixels, LAS4). In each LAS group, we primarily used the Boundary Layer Scintillometer (BLS) series scintillometer data and only used the data measured by another scintillometer (zzLAS developed by our group or Kipp & Zonen LAS) if the BLS scintillometer measurements were missing. The sites located in the three 3 × 1 and one 2 × 1 MODIS pixels were selected to study the ET upscaling methods based on the 11 EC and AWS sets, 4 LAS groups, and 3 TDP groups at sites 4–8, 11–15, and 17 (Table 1).

To estimate the differences among these flux instruments, a comparison experiment of the surface energy flux measurement systems was conducted during May 14–24, 2012, prior to conduct-

Table 1
Details of the meteorological and flux sites in three 3 × 1 and one 2 × 1 MODIS pixels.

Observation items	Type, manufactures		Height/Depth(m)	site	Duration
Eddy covariance	CSAT3/Li7500A, Campbell/Li-cor, USA		4.2 (6.2 after Aug.19)	4	site 4: 5.31—9.17
system			4.6	6	site 6: 5.28—9.21
Sensible and latent			3.8	7	site 7: 5.29—9.18
heat flux			5	13	site 13: 5.27—9.20
			4.5, 34	15	site 15: 5.25—
	CSAT3/Li7500, Campbell/Li-cor, USA		3	5	site 5: 6.03—9.18
			3.2	8	site 8: 5.28—9.21
			3.5	11	site 11: 5.29—9.18
			3.5	12	site 12: 5.28—9.21
			4.6	14	site 14: 5.30—9.21
	CSAT3/EC150, Campbell, USA		7	17	site 17: 5.31—9.17
Large aperture	BLS900, Scintec Germany		33.45	LAS1	6.07—9.19
scintillometer	zzlas, RR9340, Rainroot, China		(path length: 3256m)		6.16—9.19
Sensible heat flux					
	BLS900, Scintec Germany		33.45	LAS2	6.07—9.19
	BLS450, Scintec Germany		(path length: 2841m)		6.18—9.19
	BLS900, Scintec Germany		33.45	LAS3	6.06—9.20
	LAS, Kipp&zonen, Netherland		(path length: 3111m)		6.19—9.20
	BLS450, Scintec Germany		22.45	LAS4	6.02—
	zzlas, RR9340, Rainroot, China		(path length: 1854m)		6.21—9.20
Air Pressure	PTB110	Vaisala, Finland	–	17	AWS4: 5.10-9.17
	CS100	Campbell, USA	–	4,5,6,7,8,11,12,13,14,15,	AWS5: 6.04-9.18
Precipitation	TR-525M	Texas Electronics, USA	–	7	AWS6: 5.09-9.21
	TE525MM	Campbell, USA	–	4,5,6,8,11,12,13,14,15,	AWS7: 5.28-9.18
	52203	RM Young, USA	–	17	AWS8: 5.14-9.21
Wind speed/direction	010C/020C	Met One, USA	5,10	6	AWS11: 6.02-9.18
			10	4,5,7,8	AWS12: 5.10-9.21
	03001	RM Young, USA	10	11	AWS13: 5.06-9.20
	034B	Met One, USA	10	12,13,14,17	AWS14: 5.06-9.21
	Windsonic	Gill, UK	3,5,10,15,20,30,40	15	AWS17: 5.12-9.17;
Air	HMP45D	Vaisala, Finland	5	12,13,14	AWS15: 5.10–
temperature/humidity					
	HMP45C	Vaisala, Finland	5	4,5,17	
	HMP45AC	Vaisala, Finland	5	7,8,11	
			5,10	6	
	AV-14TH	Avalon	3,5,10,15,20,30,40	15	
Four-component	CNR4	Kipp&Zonen,	6	6,8,13,14	
radiation		Netherland	4	7,12	
	CNR1	Kipp&Zonen,	6	4, 17	
		Netherland	4	5,11	
	PSP&PIR	Eppley, USA	12	15	
Soil moisture	ECH_2O-5	Decagon Devices, USA	−0.02,−0.04,	12,13,14	
	CS616	Campbell, USA	−0.1,−0.2,−0.4,−0.6,−1;	4,5,6,7,8,11, 17	
	CS616	Campbell, USA	−0.02,−0.04,−0.1,−0.2,	15	
			−0.4,−0.8,−1.2,−1.6;		
Soil heat flux	HFP01	Hukseflux, Netherland	−0.06, −0.06, −0.06	4,5,6,7,8,11,17,	
	HFT3	Campbell,USA		12,13,14	
	HFP01SC	Hukseflux, Netherland		15	

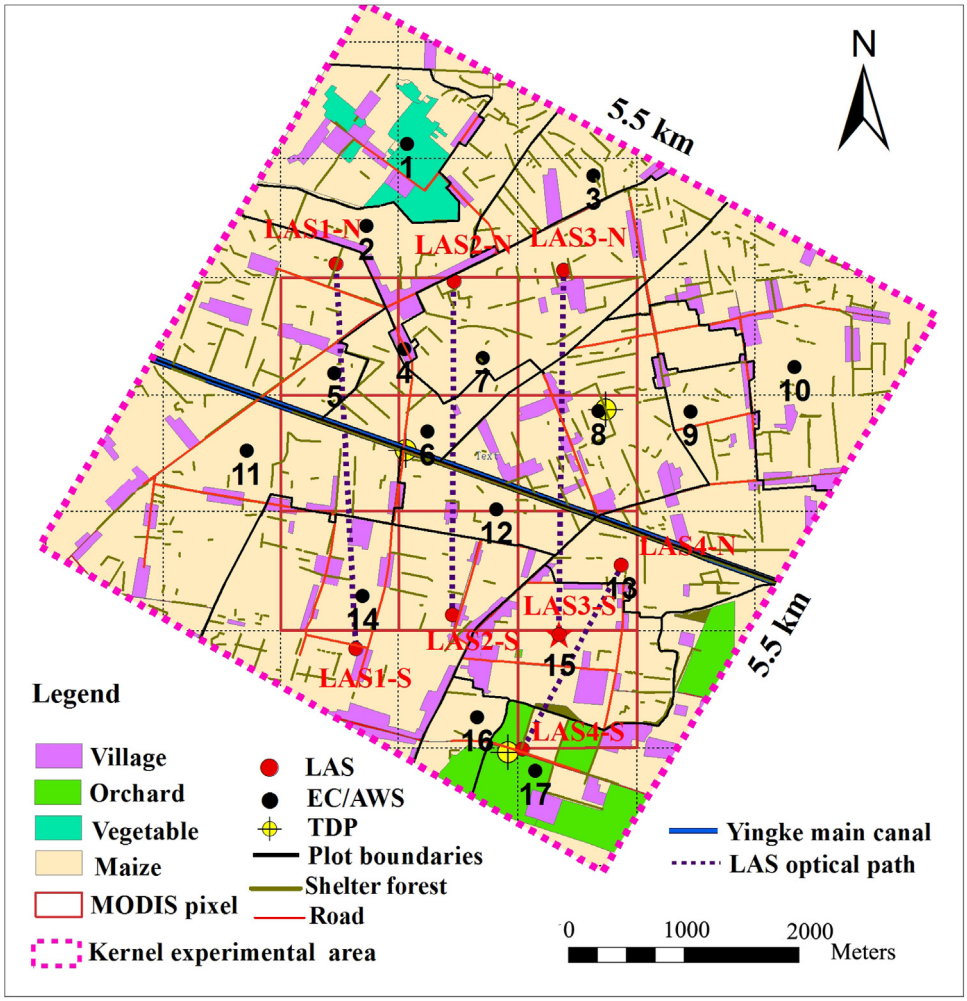

Fig. 1. Distribution of instruments in the kernel flux matrix.

ing the HiWATER-MUSOEXE. The comparison field was located in the Bajitan Gobi desert (a nearly flat and open surface), west of Zhangye city. In total, 20 EC sets, 7 LAS sets, and 18 radiometers sets involved in the comparison were used in the HiWATER-MUSOEXE. The detailed results for this experiment can found in Xu et al. (2013).

2.2. Data processing

Careful data processing and quality assessment are important to ensure the accuracy of the observation data and are critical to obtain reliable results.

2.2.1. Flux data

The raw 10 Hz EC data were processed using the Edire software developed by the University of Edinburgh (http://www.geos.ed.ac.uk/abs/research/micromet/EdiRe). The processing steps included spike detection, lag correction of H_2O/CO_2 relative to the vertical wind component, sonic virtual temperature correction, coordinate rotation (2D rotation), corrections for density fluctuations (WPL–correction), and frequency response correction. In addition, the quality of the observation data was divided into three classes according to the quality assessment method of stationarity (Δst) and the integral turbulent characteristics test (ITC): class 1 (level 0, $\Delta st < 30$, and ITC < 30), class 2 (level 1, $\Delta st < 100$, and ITC < 100), and class 3 (level 2, $\Delta st < 100$, and ITC > 100), representing high-, medium- and low-quality data, respectively. Classes 1 and 2 were selected for ET upscaling. In addition to these processing steps, the 30-min flux data were screened using a four-step procedure: (1) data from periods of sensor malfunction were rejected; (2) data within 1 h before or after precipitation were rejected; (3) incomplete 30-min data were rejected when the missing data constituted more than 3% of the 30-min raw record; and (4) data were rejected at night when the friction velocity (u_*) was below 0.1 m s^{-1} (Liu et al., 2011, 2013). To obtain daily ET, the nonlinear regression method (NLR) was used to establish the relationship between latent heat flux (EC-LE) and the net radiation (R_n) and to fill the gaps between the 30-min flux data (Berbigier et al., 2001). The daily ET was calculated using the continuous 30-min data. Due to the energy imbalance in the EC measurements, the Bowen ratio correction method proposed by Twine et al. (2000) was used to force the energy balance closure.

The LAS data were processed using the structure parameter of the air refractive index C_n^2 (measured by Scintec LAS) and the voltage $U_{C_n^2}$ ($\log C_n^2$ signal, measured by Kipp & Zonen and zzLAS). The sampling frequency was 1 Hz for the Kipp & Zonen LAS and zzLAS with an output period of 10 min, whereas the sampling frequencies were 5 Hz for BLS450 and 125 Hz for BLS 900, both of which featured an output period of 1 min. Finally, all the LAS data were averaged to 30 min. The sensible heat flux (LAS-H) was obtained via an iteration method that combines the meteorological data (e.g., wind speed, air temperature, and air pressure) based on the Monin-Obukhov similarity theory. The main quality control steps of LAS were as follows: (1) the data were rejected when C_n^2 exceeded

the saturated criterion (Ochs and Wilson, 1993); (2) the data were rejected when the demodulation signal was small; (3) the data were rejected when collected within 1 h before or after precipitation; and (4) the data were rejected if collected at night when weak turbulence occurred (u_* less than 0.1 m s^{-1}) (Liu et al., 2011, 2013). For LAS-*H* estimations during unstable atmospheric conditions, a nonlinear regression method for the relationship between LAS-*H* and the area-averaged R_n was used to estimate the missing data between the 30-min observations. During stable atmospheric conditions, the gaps were filled with a value of zero. The LAS-*LE* was estimated from the energy balance equation, and the area-averaged R_n and surface soil heat flux G_0 were obtained using the area-weighted method of related R_n and G_0 values from several plots within each LAS source area. However, during advection conditions ($H < 10$ W m^{-2} in the daytime), the LAS-*LE* was obtained directly from linear interpolation using the relationship of LAS-*LE* and the area-averaged R_n under non-advection conditions ($H > 10$ W m^{-2}).

Only daytime (10:30–18:00 BST, Beijing standard time) data under unstable conditions were used to acquire the 30-min ET data at the MODIS pixel scale. Additionally, to avoid advection conditions, we only selected 30-min data from periods when the *H* values of the EC and LAS measurements were greater than 10 W m^{-2}. The gap-filling 30-min data were only used to acquire the daily ET values at the MODIS pixel scale. In addition, days with a percentage of missing 30-min data greater than 50% were considered missing days and were not used in the upscaling analysis.

2.2.2. Remote-sensing and other auxiliary data

Land surface temperatures (LSTs) retrieved from the Advanced Spaceborne Thermal Emission and Reflection radiometer (ASTER) and the environmental protection & disaster monitoring constellation (HJ-1B) were used in this study (Zhou et al., 2015). The resolutions of ASTER and HJ-1B were 90 and 300 m, respectively, and the HJ-1B LST data were resampled to 90 m. The Normalized Difference Vegetation Index (NDVI, with a resolution of 90 m) was calculated using the reflectance in the near-infrared (ρ_{NIR}) and red bands (ρ_R). The fractional vegetation cover (FVC) was calculated using the NDVI: FVC = 1-(NDVI-$NDVI_{max}$)/($NDVI_{min}$-$NDVI_{max}$)^K, where $NDVI_{min}$ represents the normalized vegetation index of bare soil, $NDVI_{max}$ represents the maximum value of the normalized vegetation index, and *K* is 0.4631 (Baret et al., 1995). We collected 9 ASTER images (June 15, June 24, July 10, August 2, August 11, August 18, August 27, September 3, and September 12) with a passing time of 12:15 BST and 3HJ-1B images (June 17, 19, and 30) with a passing time of 12:30 BST (June 16) and 12:00 BST (June 19 and 30).

The land cover/use map was produced from a Compact Airborne Spectrographic Imager (CASI) image using the support vector machine method combined with a previous ground investigation (Zhang et al., 2013).

The automatic weather station (AWS) data were processed to 30-min average period, and data beyond the range of physical possibility were rejected. The gaps were filled via linear interpolation. The soil heat flux plates in this study were buried at depths of 0.06 m (two plates were buried under bare soil between maize plants and one plate was buried under the maize plant), and the surface soil heat flux (G_0) was calculated using the "plateCal" approach based on the weighted vegetation fraction combined with the soil temperature and moisture measured above the heat plates (Liebethal et al., 2005). The LST data were calculated from downward and upward longwave radiation measured by the AWSs and the surface emissivity measured by the FT-IR spectrometer (102F) (Mu et al., 2012). The spatial distributions of available energy, air temperature, wind speed, and air pressure from the 21 AWSs in the large experimental area were interpolated to a resolution of 90 m using the inverse distance-weighted method.

The shelterbelt transpiration was measured via a thermal dissipation probe (TDP). The raw TDP data were temperature differences with a sampling frequency of 30 s and averaged over 10 min, and processed to 30 min average period. The sap flow and sap flux density were calculated from the temperature differences and cross-sectional area of the sapwood. Then, the shelterbelt transpiration was calculated according to the shelterbelt area and tree density (Qiao et al., 2015).

3. Methods

3.1. Footprint model

The flux contributing source area of the EC and LAS measurements can be estimated using the footprint model, and the spatial representative area of the flux can be obtained quantitatively. To estimate the flux footprint, a method proposed by Kormann and Meixner (2001), which is a Eulerian analytic flux footprint model, was implemented to obtain the flux footprint of the EC measurements. The LAS source area can be calculated by combining the path-weighting function of the LAS with the footprint model for point fluxes (Meijninger et al., 2002). The footprint calculated area of EC was approximately 3 km × 3 km, with the EC installation point at the center. For the LAS footprint, the footprint calculated area was approximately 6 km × 6 km with the central path length of LAS at the center. The daily or monthly flux source area of the EC and LAS flux measurements was obtained by weighting the every half-hourly footprint results using their respective sensible heat fluxes. The resolution of the source area was 30 m for both EC and LAS measurements, and the flux contribution of the chosen total source area was set to 90%.

3.2. Upscaling methods

3.2.1. Three simple upscaling methods

Three simple upscaling methods for ET were used in this study: the arithmetic average, area-weighted average, and footprint-weighted methods.

The arithmetic average method was primarily used for homogeneous land surfaces, where each plot plays the same role and the pixel values can be acquired by averaging several EC measurements.

For the area-weighted and footprint-weighted methods, the study area was first divided into different plots according to the surface heterogeneity; then, the fraction of each plot or the relative weight of the LAS footprint at the satellite pixel scale was acquired combining the land use/cover map and footprint model; and finally, the latent heat flux at the satellite pixel scale was computed using the following relationship (Eq. (1)).

$$LE_{scaling} = \sum_{i=1}^{N} W_i \theta_i \quad (1)$$

where $LE_{scaling}$ is the latent heat flux in the three 3 × 1 or one 2 × 1 MODIS pixels, *N* is the number of plots in the pixel, W_i is the area proportion of the *i* th plot relative to the satellite pixel or the relative weight of the LAS footprint at the satellite pixel scale (Table 2), and θ_i is the EC measurement in the *i* th plot. The relevant ECs are as follows: LAS1: EC4-6, EC11, and EC14; LAS2: EC4, 6, 7, 8, 12, and 14; LAS3: EC4, 7, 8, 12, 13, and 15; and LAS4: EC4, 12, 13, 15, and 17 (Fig. 1).

Table 2
The weighted value of each EC measurement in the different MODIS pixels.

Related EC	3 × 1MODIS pixels (LAS1)		3 × 1MODIS pixels (LAS2)		3 × 1MODIS pixels (LAS3)		2 × 1MODIS pixels (LAS4)	
	Area weighted (%)	Footprint weighted (%)	Area weighted (%)	Footprint weighted (%)	Area weighted (%)	Footprint weighted (%)	Area weighted (%)	Footprint weighted (%)
EC4	15	7	15	10	13	31	7	10
EC5	20	19	–	–	–	–		
EC6	13	22	17	18	–	–		
EC7	–	–	19	17	12	6		
EC8	–	–	9	18	36	31		
EC11	19	33	–	–	–	–		
EC12	–	–	14	18	10	15	10	2
EC13	–	–	–	–	9	8	21	7
EC14	22	14	19	13	–	–		
EC15	–	–	–	–	11	5	50	74
EC17	–	–	–	–	–	–	6	5
shelterbelt	11	5	7	6	9	4	6	2

3.2.2. The integrated Priestley-Taylor equation method

The Priestley-Taylor equation (Priestley and Taylor, 1972) can be used to calculate the potential ET from wet surfaces under conditions of minimal advection as follows:

$$LE_p = \alpha \frac{\Delta}{\Delta + \gamma}(R_n - G_0) \quad (2)$$

where LE_p is the latent heat flux for saturated surfaces, α is the Priestley-Taylor parameter, R_n and G_0 are the net radiation and surface soil heat flux, respectively, Δ is the slope of the saturation vapor pressure to air temperature, and γ is the psychrometer constant.

Multiple researchers have estimated the actual ET over unsaturated surfaces using the Priestley-Taylor equation by establishing a relationship between α and the soil moisture, land surface temperature, and vegetation parameters (Davies and Allen, 1973; Jiang and Islam, 2001; Fisher et al., 2008).

In this study, ET at the satellite pixel scale was acquired using the Priestley-Taylor equation by combining the land surface temperatures retrieved from remote sensing images. The processing steps were as follows: 1) The α for each EC site was calculated using the Priestley-Taylor equation by combining the observations of LE measured by EC and R_n, G_0, the air temperature (T_a), and the air pressure (P) from the AWS in each plot. 2) The α values were plotted versus their influencing factors, such as the soil moisture, the difference between the surface temperature and air temperature (T_s-T_a), and the vapor pressure deficit (VPD), to determine the most influential factors. Then, the relationship between α and the most influential factor was established in each plot. This relationship was considered suitable for the entire plot, and the α in each plot was estimated from AWS and remote sensing data. 3) The ET in each plot was calculated using the Priestley-Taylor equation and the interpolated air temperature, air pressure and available energy. Finally, ET estimates at the satellite pixel scale were acquired using Eq. (3), which combines the measurements from the residential area (EC4) and the shelterbelts (around sites 6, 8, and 17).

$$LE_{scaling} = W_i LE_i + W_{village} LE_{village} + W_{shelterbelt} LE_{shelterbelt} \quad (3)$$

where $LE_{scaling}$ is the latent heat flux in the three 3 × 1 or one 2 × 1 MODIS pixels, LE_i is the latent heat flux in the i th plot calculated by the Priestley-Taylor equation, $LE_{village}$ is the latent heat flux measured in the residential area, $LE_{shelterbelt}$ is the latent heat flux measured by TDP in the shelterbelt, W_i is the fraction of the ith plot relative to the three 3 × 1 or one 2 × 1 MODIS pixels, and $W_{village}$ and $W_{shelterbelt}$ are the fractions of residential area and shelterbelt relative to the three 3 × 1 or one 2 × 1 MODIS pixels, respectively.

3.2.3. The area-to-area regression kriging method

Currently, the geostatistics method is rarely used in upscaling surface fluxes. However, with the rapid development of geostatistics in recent years, the area-to-area regression kriging method provides a new way to upscale surface fluxes under heterogeneous land surfaces (Ge et al., 2015). The sensible heat flux of the three 3 × 1 or one 2 × 1 MODIS pixels was obtained using the area-to-area regression kriging method by combining the EC measurements made during the HiWATER-MUSOEXE and the FVC, LST, and NDVI data retrieved from remote sensing images. First, a linear multiple regression model was established to relate the auxiliary variables for the surface sensible heat flux trend and residual:

$$H_{it} = \beta_{0t} + \beta_{1t} \cdot \overline{FVC_{it}} + \beta_{2t} \cdot \overline{LST_{it}} + \beta_{3t} \cdot \overline{NDVI_{it}} + \beta_{4t} \cdot \overline{u_{it}} + r_{it} \quad (4)$$

where H_{it} is the sensible heat flux of the i th EC during the t th period;$\overline{FVC_{it}}$, $\overline{LST_{it}}$, $\overline{NDVI_{it}}$, and $\overline{u_{it}}$ are auxiliary environmental variables, i.e., the FVC, LST, NDVI, and wind speed (u), extracted by the i th EC footprints during the t th period; β_{it} (0, 1, 2,. . .) are regression parameters; and r_{it} is the residual of the i th EC during the t th period.

Second, the regression coefficient was achieved by the stepwise regression method, namely, β_{it} (0, 1, 2,. . .) and r_{it}. Third, the residual of the area-averaged sensible heat flux was calculated using the area-to-area regression kriging method, and then, the sensible heat flux was estimated by adding the trend (regression estimations) to the area-to-area regression kriging results (residual at the three 3 × 1 or one 2 × 1 MODIS pixels). Finally, the latent heat flux at the three 3 × 1 or one 2 × 1 MODIS pixels was obtained using the residual method of the surface energy balance equation. Details of this method can be found in Ge et al. (2015).

3.3. Calculation of the area overlap degree and spatial heterogeneity

Based on the footprint model and LST retrieved from ASTER, Ave_RW and ΔT were used to characterize the degree of overlap between the source areas of the EC and LAS3 pixels, the spatial heterogeneity of the surface water and heat conditions in the LAS3 pixels, respectively. The equations used are as follows:

$$Ave\_RW = \frac{\sum_{i=1}^{n} area(i)_{\_EC}}{area_{\_pixel}} \quad (5)$$

$$\Delta T = |\sum_{j=1}^{n} (\overline{T}_{s(j)\_EC} - \overline{T}_{s(j)\_plot}) f_{j\_plot}| \quad (6)$$

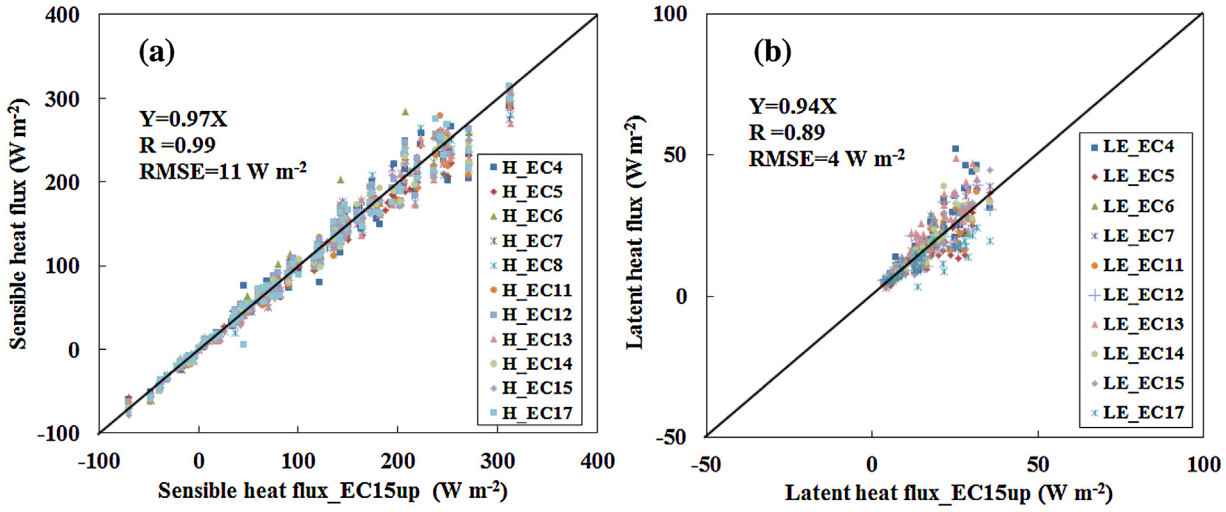

Fig. 2. Comparison of sensible and latent heat fluxes measured by EC (using the measurements of EC15-up in HiWATER-MUSOEXE as the reference, May 14–24, 2012).

where n is the number of plots in the LAS3 pixels; area (i)$_{\_EC}$ and area$_{_pixel}$ are the area of the EC source area in the *i* th plot and the LAS3 pixels, respectively; ΔT is the surface temperature difference between the LAS3 pixels and the EC source area; $\sum_{j=1}^{n}\overline{T}_{s(j)\_EC}$ is the mean surface temperature within the EC source area in the *j* th plot; $\bar{T}_{s(j)\_plot}$ is the mean surface temperature in the *j* th plot; and f_{j_plot} is the area proportion of the *j* th plot relative to the LAS3 pixels.

3.4. The uncertainty assessment methods

To assess the uncertainty of the upscaled ET results, the indices of MAPE, relative error (RE), root mean square error (RMSE), and correlation coefficient (R) were selected, and their respective equations are as follows:

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{|P_i - O_i|}{\bar{O}} \tag{7}$$

$$RE = \frac{P_i - O_i}{O_i} \tag{8}$$

$$RMSE = \sqrt{\sum_{i=1}^{n}(P_i - O_i)^2/n} \tag{9}$$

$$R = \sum_{i=1}^{n}\left(P_i - \bar{P}\right)\left(O_i - \bar{O}\right) / \left[\sum_{i=1}^{n}\left(P_i - \bar{P}\right)^2\sum_{i=1}^{n}\left(O_i - \bar{O}\right)^2\right]^{1/2} \tag{10}$$

where P_i is the estimated value, O_i is the measured value, $\bar{P}$ is the mean estimated value, $\bar{O}$is the mean measured value, and n is the number of samples. In Section 4.1, *P* indicates measurements of the same type instruments, and *O* indicates measurements of the reference instrument or the LAS measurements. In Section 4.3 and 4.4, *P* indicates the upscaling values, and *O* indicates the LAS measurements. The uncertainties of the upscaled ET results primarily originate in the ground observations (EC and LAS) and upscaling process.

4. Results and discussion

4.1. Consistency comparison of the surface energy flux instruments

The results of the ECs, LASs, and radiometers were compared from May 14 to May 24, 2012, in the Bajitan Gobi desert before the HiWATER-MUSOEXE was conducted. In this section, the 11 EC sets and 4 LAS sets (LAS1-3: BLS900; LAS4: BLS450) used in this study were chosen to analyze the instrumental consistency. The EC installed at site 15 (up layer) (CSAT3 and Li7500A) and the LAS in LAS1 pixels (BLS900) during the HiWATER-MUSOEXE were selected as the references, and the sensible and latent heat fluxes measured by EC and LAS were compared within the three 3×1 and one 2×1 MODIS pixels (Figs. 2 and 3). The EC8 was recalibrated during the comparison period due to its low H_2O density, and the latent heat flux is not plotted in Fig. 2. Both the sensible heat flux and latent heat flux measured by EC corresponded well; the RMSE, regression slope and R of the sensible heat flux and latent heat flux were 11 W m^{-2} and 4 W m^{-2}, 3% and 6%, and 0.99 and 0.89, respectively. A degree of scatter is observed in Fig. 2b due to the lower latent heat flux at the Bajitan Gobi station (less than 50 W m^{-2}). The LAS measurements also corresponded well with each other taking the measurements of LAS1 as the reference (R = 0.99, Fig. 3) and featured RMSE and regression slope values of 10 W m^{-2} and 2%, respectively.

The LAS measurements (LAS1-LAS4) are compared with the related EC measurements in Fig. 4. As depicted in Fig. 4, all LAS groups and related EC measurements were consistent with each other. The RMSE values of LAS1, LAS2, LAS3 and LAS4 were 21 W m^{-2}, 16 W m^{-2}, 14 W m^{-2}, and 17 W m^{-2}; the regression slopes were 3%, 1%, 1%, and 5%; and the R values were 0.97, 0.98, 0.98, and 0.97, respectively.

The above analysis shows that both the EC measurements and LAS measurements were consistent with each other for a homogeneous land surface and that the LAS measurements and related ECs were also comparable.

4.2. Temporal-spatial variations of ET in the kernel experimental area

The ET process is influenced by meteorological conditions, vegetation status, soil moisture, and topography, etc. The study area was located in an oasis-desert ecosystem, and the surfaces of the oasis varied among cropland, orchards, residential areas, roads and

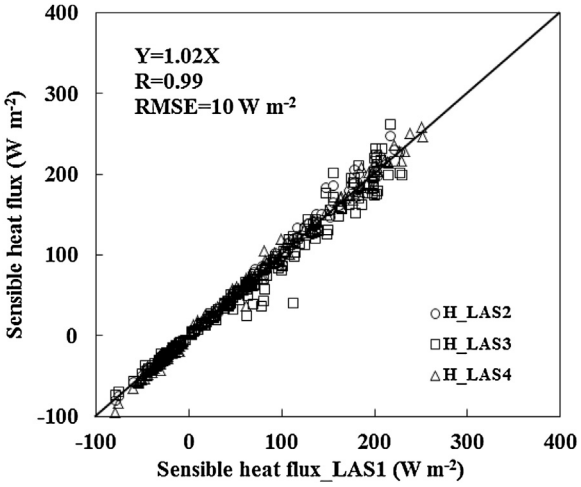

Fig. 3. Comparison of sensible heat fluxes measured by LAS (using the measurements of LAS1 in HiWATER-MUSOEXE as the reference, May 14–24, 2012).

shelterbelts. The thermal and dynamic characteristics of these surfaces were different. In the oasis, surface water and heat conditions were also different, even in the same cropland, due to different soil moisture contents (associated with different irrigation times and quantities) and different crop growing conditions. In addition, the different densities, heights, and orientations of the shelterbelts caused differences in the surface roughness. In brief, the shelterbelts and different surfaces induced dynamic heterogeneity, and the different irrigation patterns and land surfaces induced thermal heterogeneity. Together, these factors caused spatial ET heterogeneity. During the crop growing period, the surface heterogeneity changed over time, leading to corresponding changes in ET.

The soil moisture, plant growth status, and meteorological elements (i.e., radiation, precipitation, and wind speed, etc.) differed among all plots within the LAS1, LAS2, LAS3 and LAS4 pixels. Therefore, the EC measurements in each plot were different, and the LAS measurements which represented the satellite pixel values (three 3×1 and one 2×1 MODIS pixels), were also different. These differences also changed over time during the crop growing period. The seasonal variations in latent heat flux during 10:30–18:00 BST are shown in Fig. 5.

The amplitude of variation in the latent heat flux was smallest for the residential surface (site 4) because the surface characteristics varied only slightly over time (Fig. 5). In contrast, the EC measurements at other sites (i.e., cropland (sites 5–8, 11–15) and orchard (site 17)) and the LAS measurements (LAS1-4) increased with plant growth and gradually decreased at the end of the growing period. In general, the LAS measurement values were within those of the EC measurements for the cropland, orchard and residential surfaces. The LAS3 measurements differed considerably from these EC measurements, while the differences between LAS1 and related EC measurements were relatively small. In general, the latent heat flux in the orchard was greater than that in the cropland, and the smallest latent heat flux was in residential areas. The latent heat flux in each site increased after irrigation or pre-

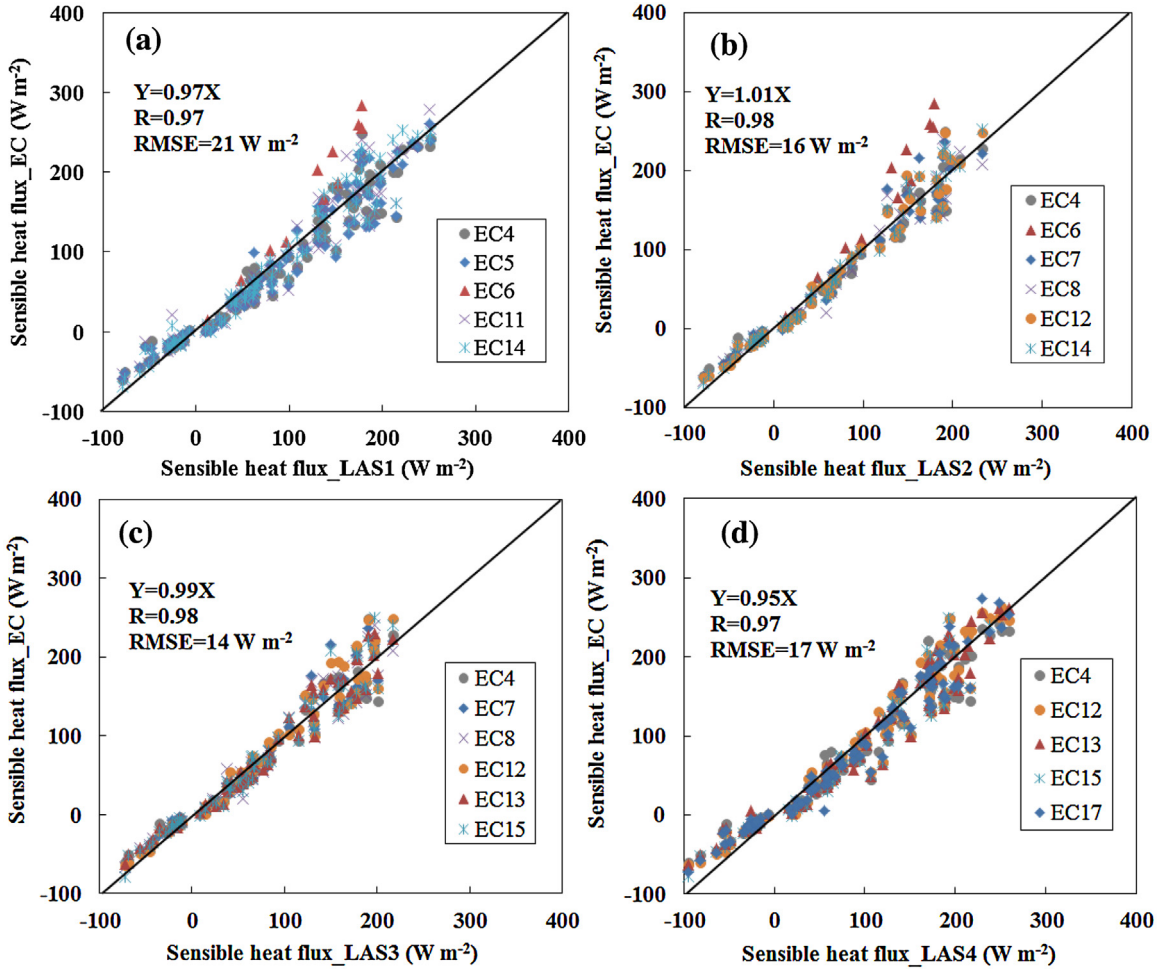

Fig. 4. Comparison of sensible heat fluxes measured by LAS and associated EC systems: (a) LAS1, (b) LAS2, (c) LAS3 and (d) LAS4 (May 14–24, 2012).

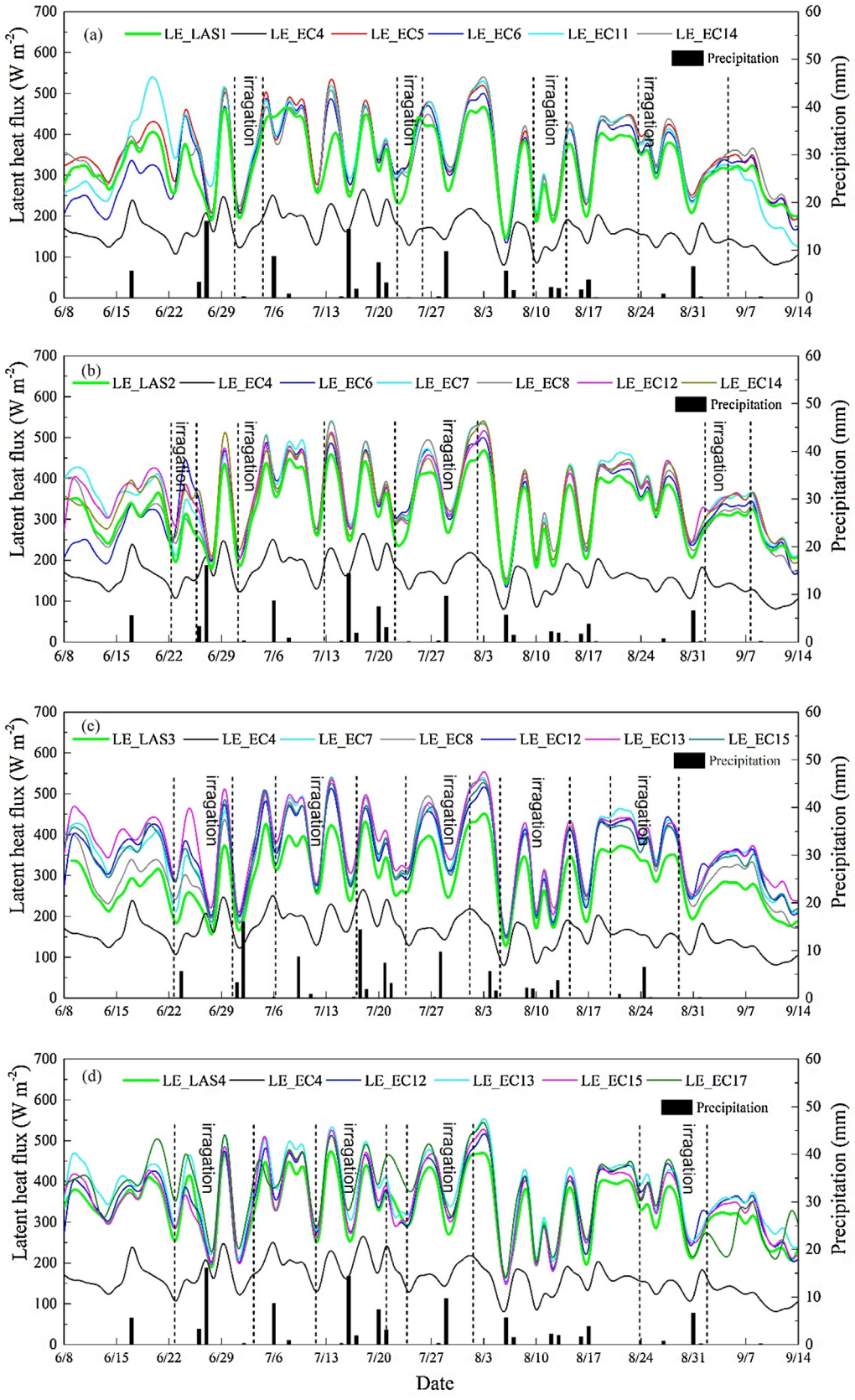

Fig. 5. The seasonal variation in the latent heat flux measured by LAS and related EC groups in the kernel experimental area between June 8 and September 14, 2012. (a), (b), (c) and (d) correspond to LAS1, LAS2, LAS3 and LAS4, respectively, averaged across 10:30–18:00 BST.

cipitation, and the latent heat flux among each site also differed considerably during the irrigation periods.

The above analysis shows that a single EC measurement cannot represent the satellite pixel values, i.e., within the LAS1, LAS2, LAS3, and LAS4 pixels (three 3 × 1 and one 2 × 1 MODIS pixels) and that the differences between the EC and LAS measurements changed over time.

4.3. Upscaling multi-site ET measurements using three simple upscaling methods

The flux observation matrix data from June 8 to September 14, 2012, were selected for the analysis combing the land use/cover map and footprint model, including the 11 EC sets (EC4-EC8, EC11-EC15, EC17), 4 group LAS sets (LAS1-LAS4), and 3 TDP sets in the three 3 × 1 and one 2 × 1 MODIS pixels (Fig. 1).

Each EC system in the HiWATER-MUSOEXE was installed on a relatively flat, homogeneous land surface within each plot, and the measurement height was approximately 3–5 m (EC17: 7 m; the height of EC4 was adjusted to 6.2 m after August 19). If the surface water and heat conditions at each plot are assumed to be relatively uniform, these EC measurements can be considered to represent the average conditions of each plot. Each group of LAS sets covers 3 × 1 or 2 × 1 MODIS pixels, and the effective heights were approximately 35 m (LAS1-3) or 23 m (LAS4), which are greater than the shelterbelt heights (approximately 12.5–25 m in the oasis). The four groups of LAS sets were installed above the internal boundary layer; therefore, the LAS measurements were considered to represent the area-averaged water and heat fluxes in the three 3 × 1 or one 2 × 1 MODIS pixels. The instantaneous footprint (ASTER passing time: 12:30 BST), the daily footprint on June 15, and the climatology footprints between June 8 and September 14 (10:30–18:00 BST) were calculated (Fig. 6).

As shown in Fig. 6, the proportion of instantaneous LAS source area in the 3 × 1 or 2 × 1 MODIS pixels was greater than 50%. The source area was located in the central area and covered all the corresponding plots (Fig. 6a). The daily source area covered nearly all the 3 × 1 or 2 × 1 MODIS pixels with a proportion greater than 80% (Fig. 6b). The proportion of daytime averaged source area during the entire growing period was also greater than 80%. The LAS source areas (LAS1-LAS3) were typically approximately 3000 m long and 1000 m wide, and the LAS4 source area was approximately 2000 m × 1000 m. Therefore, the LAS source areas were considered to cover the 3 × 1 or 2 × 1 MODIS pixels, and the LAS measurements can represent area-averaged values in the 3 × 1 or 2 × 1 MODIS pixels. The LAS measurements could be taken as a reference to evaluate the upscaled results in the 3 × 1 or 2 × 1 MODIS pixels.

The simple upscaling methods, including the arithmetic average, area-weighted, and footprint-weighted methods, were used to upscale the ET observations from the multi-site EC measurements. Then, the “ground-truth” ET could be acquired at the satellite pixel scale. The weights of the EC sites in the area-weighted and footprint-weighted methods are shown in Table 2.

Among the upscaled ET results based on the three simple upscaling methods, the results obtained with the arithmetic average method displayed relatively large scatter along the 1:1 line, compared to the LAS observations (Fig. 7). Therefore, the area-weighted and footprint-weighted methods were the focus of the following analysis.

The upscaled results obtained from the area- and footprint-weighted methods were similar in the LAS1, LAS2, and LAS4 pixels. For the latent heat flux results for the LAS1 pixels using the area- and footprint-weighted methods, the MAPE values were 7.7% and 5.3%, the RMSE values were 32 W m^{-2} and 23 W m^{-2}, and the R values were 0.99 and 0.98, respectively. For the LAS2 pixels, the MAPE values were 5.4% and 4.3%, the RMSE values were 22 W m^{-2} and 17 W m^{-2}, and the R values were 0.98 and 0.99, respectively. For the LAS4 pixels, the MAPE values were 5.3% and 4.5%, the RMSE values were 25 W m^{-2} and 22 W m^{-2}, and the R values were 0.97 and 0.98, respectively. The upscaled results of latent heat flux in the LAS3 pixels were relatively poor, (MAPE values of 10.8% and 7.6%, RMSE values of 37 W m^{-2} and 27 W m^{-2}, and relative low R values of 0.96 and 0.95) due to the relatively high heterogeneity in the LAS3 pixels. From the above analysis, we see that the upscaled results of the simple averaging methods were affected by the heterogeneity of the underlying surfaces. In the subsequent analysis, the area-weighted method was used to analyze the variations in the upscaled results with time (Fig. 8).

As shown in Fig. 8, the REs of the LAS1, 2, and 4 pixels during the crop growing period were approximately −15% to 8%, −10% to 11%, and −10% to 10%, respectively, with MAPE values of 7.7%, 5.4% and 5.3%, respectively. The upscaled results from the area-weighted method were reliable. However, the REs of the LAS3 pixels were relatively high (−10% to 30%) and exhibited a U-shaped variation (Fig. 8c), with a MAPE value of 10.8% during the growing period. The REs were high before full cover (before July 10; REs between −14.0% and 52.3% with a mean MAPE of 16.3%) and after the end of the growing period (after September 4; REs between −17.8% and 52.9% with a mean MAPE of 10.0%). In the middle of the crop growing period (from July 10 to September 4), the REs were relatively low (−12.4% to 28.2%), with a mean MAPE of 6.6%. This phenomenon can be attributed to the heterogeneity of the underlying surfaces, i.e., the influence of irrigation and different crop growing conditions. The surface was highly heterogeneous before full cover (July 10) because the maize was interspersed with bare soil. Therefore, the assumption that the EC observations represent the area-averaged ET of each plot was not entirely appropriate. The surface was primarily covered by maize after full cover, and the heterogeneity of the underlying surfaces was reduced greatly, except during irrigation periods, when some fields were irrigated and other fields were not. At the end of the crop growing period, the heterogeneity of the underlying surfaces increased due to frost-withered maize. Therefore, the surface heterogeneity determined the validity of the assumption that the EC measurements represent the area-averaged ET of each plot, and produced an effect on the upscaled results.

Liu et al. (2011) considered that the differences between the sensible heat fluxes derived from LAS and EC were caused by the differences between the source areas of the EC and LAS measurements, the heterogeneity of the underlying surfaces in their source areas and the EC energy imbalance. Hoedjes et al. (2007) found that LSTs obtained from thermal infrared satellite imagery can provide a good indication of the degree of spatial heterogeneity based on water and heat conditions. In this study, the LSTs from nine ASTER images between May and September 2012 were used to further analyze the influences of the heterogeneity of underlying surfaces on the upscaled results for the LAS3 pixels.

The relationship among the MAPE values of the latent heat flux upscaling results (Eq. (7)), the degree of overlap between the source areas of the EC and LAS3 pixels (Ave_RW) (Eq. (5)), and the spatial heterogeneity of the surface water and heat conditions in the LAS3 pixels (ΔT) (Eq. (6)) were obtained. As shown in Fig. 9, Ave_RW varied only slightly during the crop growing period, with a value of approximately 9%. Therefore, the degree of overlap between the source areas of EC and the LAS3 pixels does not significantly influence the MAPE variation in the upscaled results. The variations in the upscaled results were consistent with those of ΔT, which characterized the heterogeneity of surface water and heat conditions in the LAS3 pixels. The heterogeneity of the underlying surfaces was high before full cover (i.e., ΔT values of 1.9 K on June 15 and 2.1 K on June 24), and the upscaled results were relatively poor (i.e., MAPE values of 27.2% on June 15 and 42.1% on June 24). In the middle of the

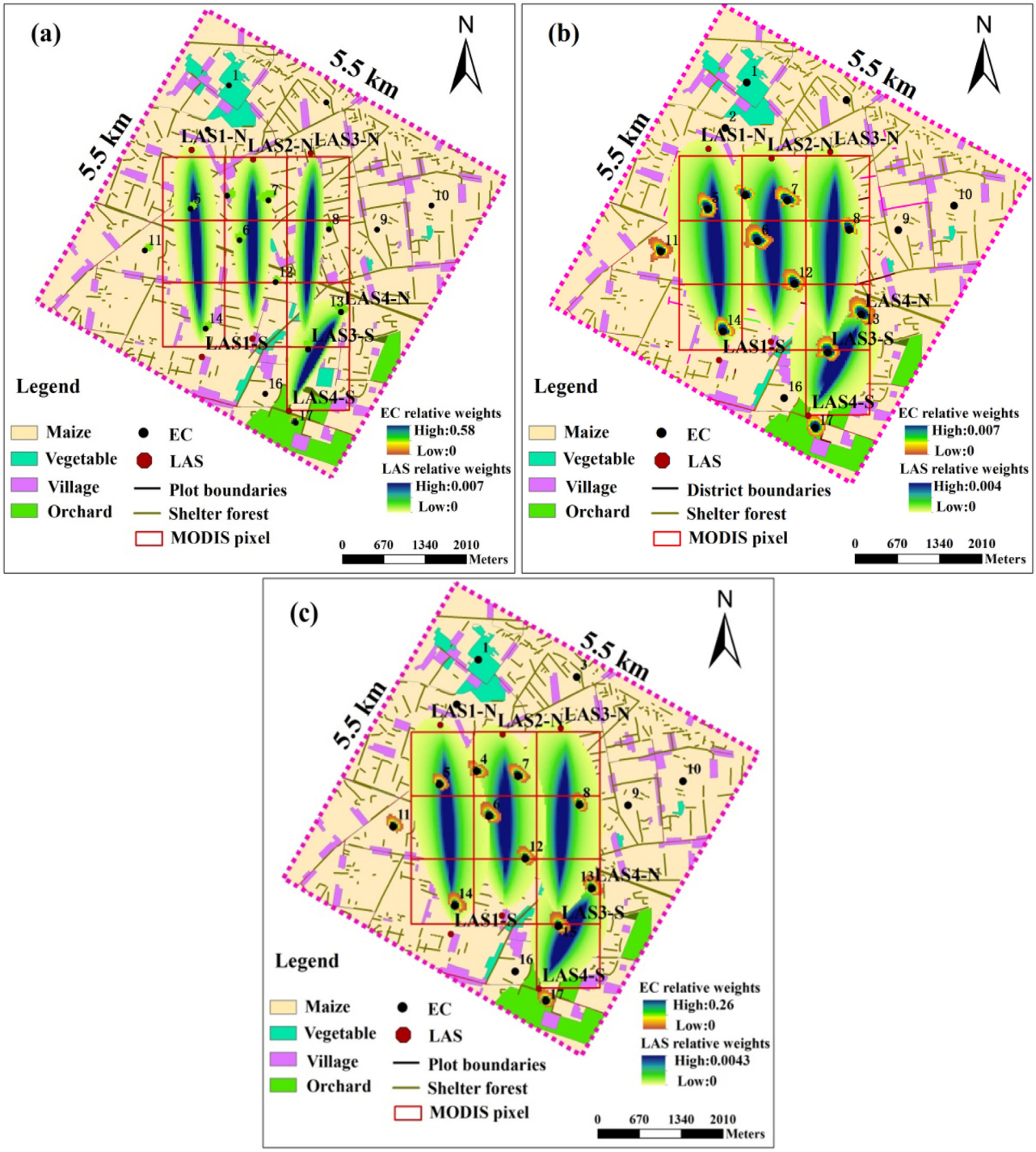

Fig. 6. The LAS and EC source areas (90% flux contribution): (a) ASTER passing time on June 15, (b) ASTER passing day on June 15 and (c) daytime averaged source area between June 8 and September 14.

crop growing period (from July 10 to September 4), the surface was relatively homogenous, and the upscaled results were good (MAPE and ΔT values of less than 10% and 1.5 K, respectively). At the end of the crop growing period (after September 4), the heterogeneity of the underlying surfaces increased due to frost-withered maize, and the accuracy of the upscaled results decreased (i.e., MAPE and ΔT values of 23.3% and 1.6 K, respectively, on September 12). When irrigation occurred in the middle of the crop growing period (three times between July 10 and September 4), the heterogeneity of the underlying surfaces and therefore the MAPE values of the upscaled results increased because only some of the plots were irrigated (e.g., on August 11, the plots of EC7 and EC12 within the LAS3 pixels were irrigated, but the entire area was not). The MAPE and ΔT values increased to the values of 11% and 1.51 K, respectively.

Due to the high levels of heterogeneity in the underlying surfaces of the LAS3 pixels, the MAPE was large before full cover and at the end of the crop growing period, and the variations in the MAPE were consistent with those of ΔT. Therefore, the simple upscaling methods could not achieve satisfactorily upscaled ET results over heterogeneous land surfaces, and auxiliary variables that can

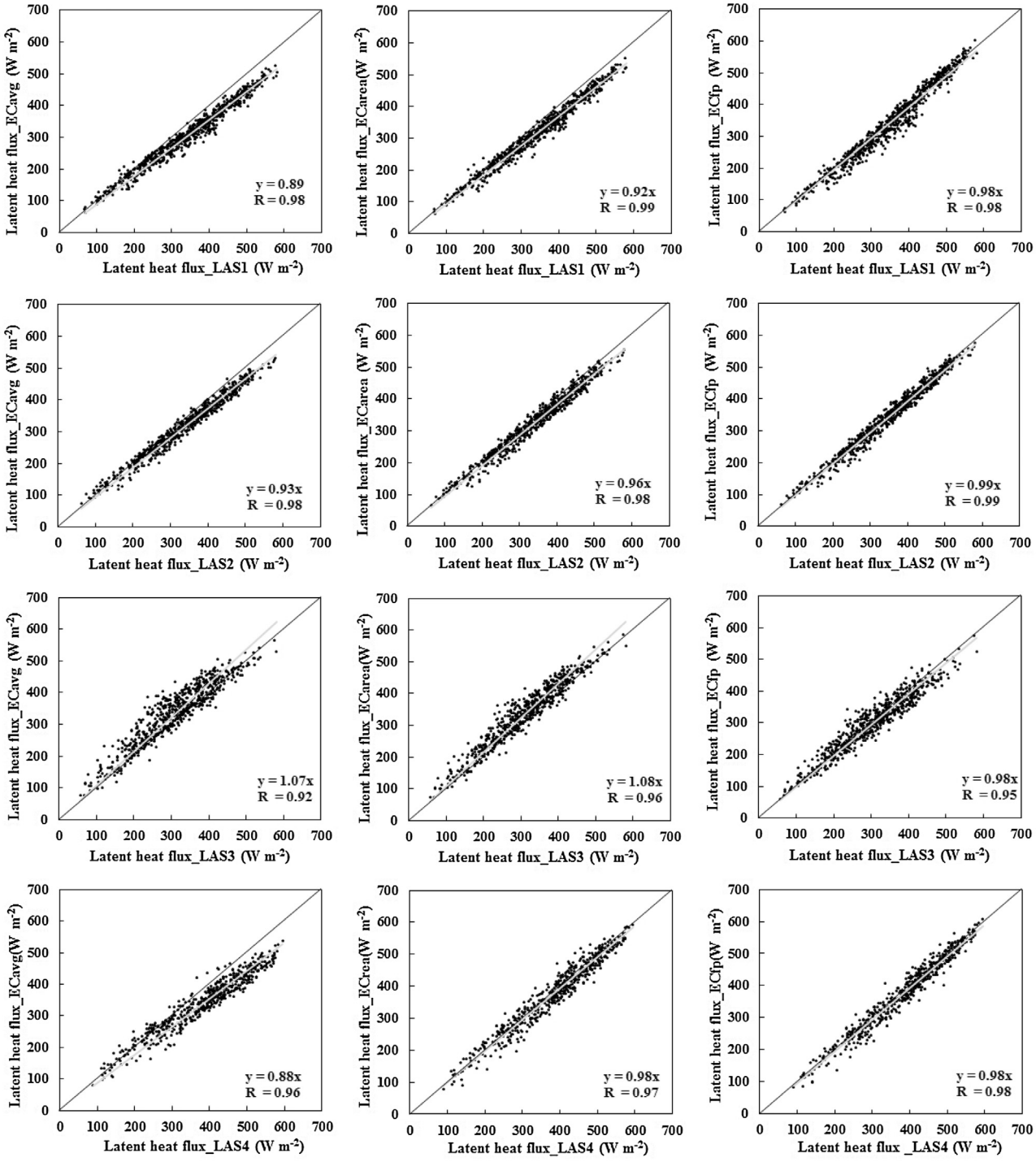

Fig. 7. Comparison of the results from three simple upscaling methods between June 8 and September 14, 2012: (left) the arithmetic average method, (center) the area-weighted method and (right) the footprint-weighted method; LAS1-LAS4 from top to bottom.

characterize the spatial heterogeneity of surface water and heat conditions must be introduced.

4.4. Upscaling multi-site ET measurements using upscaling methods introducing auxiliary variables

The measurements of a single EC set cannot represent area-averaged ET for each plot with heterogeneous land surfaces, and the accuracy of the upscaled results was reduced when the heterogeneity of surface water and heat conditions increased. Therefore, land surface temperatures that can reflect surface water and heat conditions were introduced, and multi-site ET measurements were upscaled to estimate ET at the satellite pixel scale using the upscaling method introducing auxiliary variables, such as the integrated Priestley-Taylor equation method or area-to-area regression kriging method.

With the latent heat flux dataset based on EC and AWS measurements (e.g., net radiation, surface soil heat flux, air temperature, pressure, and LST) in 11 plots within the three 3×1 and one 2×1 MODIS pixels, the Priestley-Taylor parameter α can be deduced

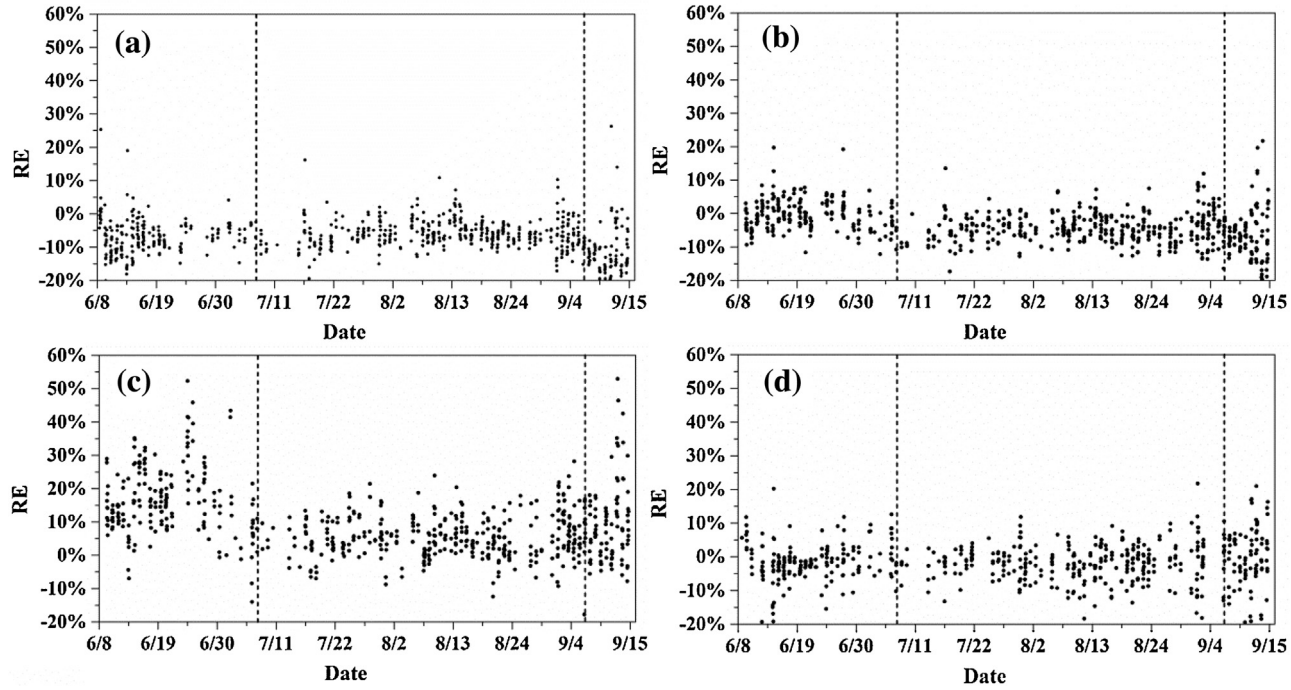

Fig. 8. Change in the relative error of the upscaling latent heat flux derived from EC using the area weighted method. (a), (b), (c) and (d) represent LAS1, LAS2, LAS3, and LAS4, respectively; the two dot lines represent the crop at full cover and at the end of the crop growing period, respectively.

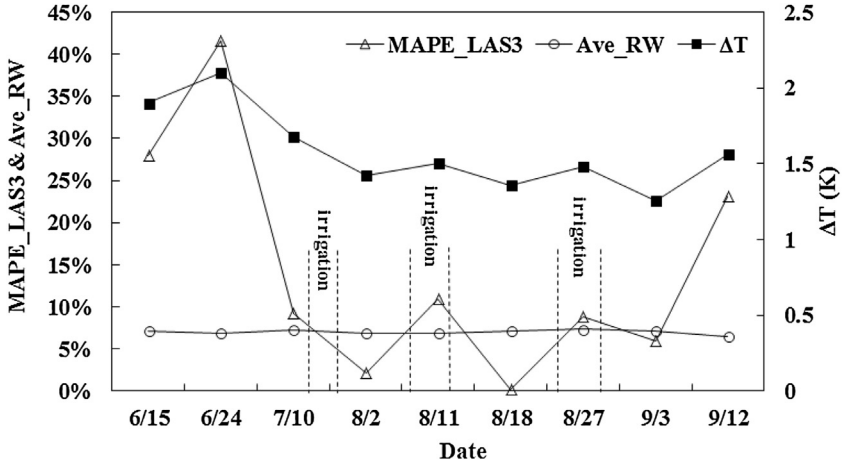

Fig. 9. Relationship between the MAPE values of the upscaled latent heat flux results, the degree of overlap between source areas of the EC and LAS3 pixels (Ave_RW) and the heterogeneity of the surface water and heat conditions in the LAS3 pixels (ΔT) between June 8 and September 14, 2012.

Table 3
Relationship ($y = a \cdot x + b$) between the Priestley-Taylor parameter α (y-axis) and differences between the surface and air temperatures, T_s-T_a (x-axis) in each plot.

Plot	June 8 to July 9				September 5 to September 14			
	a	b	R	n	a	b	R	n
EC5	−0.036	1.249	0.940	40	−0.078	1.152	0.804	20
EC6	−0.026	1.216	0.936	56	−0.108	1.233	0.820	20
EC7	−0.009	1.237	0.685	47	−0.062	1.057	0.841	21
EC8	−0.032	1.248	0.849	52	−0.087	1.141	0.867	18
EC11	−0.025	1.231	0.901	56	−0.101	1.089	0.850	19
EC12	−0.028	1.266	0.903	55	−0.071	1.192	0.802	20
EC13	−0.014	1.351	0.820	53	−0.055	1.339	0.759	18
EC14	−0.032	1.259	0.940	56	−0.080	1.203	0.852	20
EC15	−0.030	1.270	0.837	56	−0.070	1.205	0.829	20
EC17	−0.529	1.244	0.846	54	−0.070	1.702	0.817	20

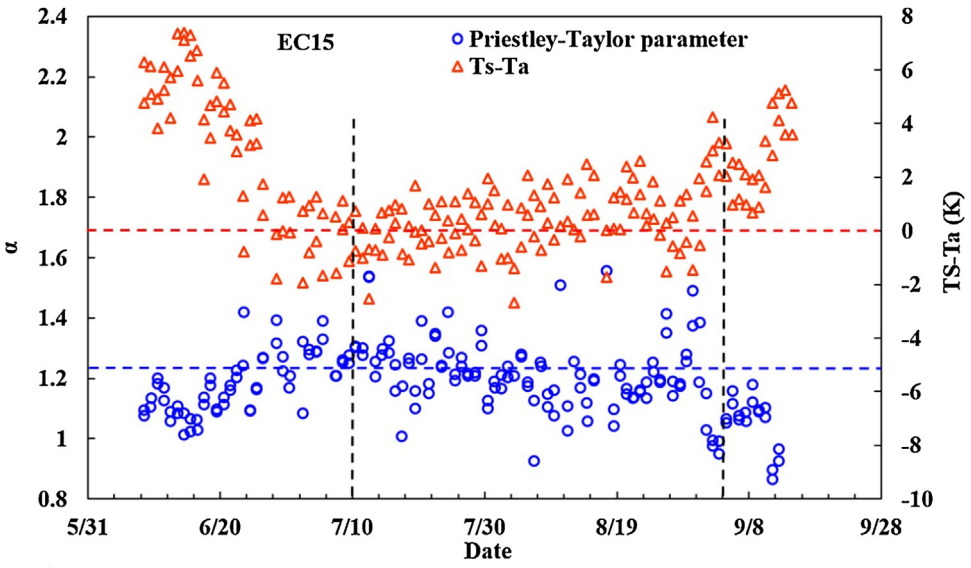

Fig. 10. Variation in the Priestley-Taylor parameter α and the difference between the surface and air temperatures (T_s-T_a) between June 8 and September 14, 2012 (EC15). The black dotted lines indicate the crop at full cover and at the end of the crop growing period, respectively, the blue dotted line indicates α = 1.26 and the red dotted line indicates T_s-T_a = 0 K (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

using the Priestley-Taylor equation and ground measurements. Then, the relationship between α and its influencing factors can be established. Due to the four irrigation events, the soil moisture remained at a high value (the volumetric water content was approximately 30%) in the oasis cropland, and the vapor pressure deficit (VPD) was small. Therefore, no significant relationship between α and soil moisture or VPD existed, rather the difference between the land surface temperature and air temperature (T_s-T_a) was the primary factor influencing ET in the oasis cropland. Fig. 10 shows the variation of α and T_s-T_a during the crop growing period (using EC15 as an example). As shown in Fig. 10, a clear negative correlation exists between α and T_s-T_a before full cover (July 10) and at the end of the growing period (after September 4). In contrast, α was approximately equal to the historical value of 1.26 during the middle of the crop growing period (blue dotted line) because the cropland was fully covered by maize and the soil moisture was nearly at field capacity (i.e., a volumetric water content of approximately 20–40% in the depth of 0–80 cm). The LSTs retrieved from remote sensing images were used in the following upscaling process; therefore, the relationship between α and T_s-T_a was estimated for the satellite passing times (11:30-12:00 BST and 12:00-12:30, BST). The relationship between α and T_s-T_a was established using a linear equation in each plot before full cover and at the end of the crop growing period (Table 3, the residential plot (EC4) was applied directly in Eq. (3), therefore EC4 was not shown in the table). A good correlation coefficient was achieved, and the results were significant at the 0.01 level.

The heterogeneity of the surface water and heat conditions was high before full cover (July 10) and at the end of the crop growing period (after September 4). This heterogeneity produced relatively large errors in the upscaled ET results using the simple upscaling methods. In the following analysis, T_s-T_a was first calculated based on the LSTs retrieved from ASTER (June 15, June 24, and September 12; passing time: 12:30 BST), HJ (June 16, June 19, and June 30; passing time: 12:30 BST) and air temperature interpolated by the inverse distance-weighted method. Then, the Priestley-Taylor parameter α was derived from the relationships in Table 3, and the latent heat flux at each plot in the cropland was calculated based on the Priestley-Taylor equation. Finally, the latent heat flux in the LAS3 pixels was acquired using Eq. (3). The comparisons between the upscaled multi-site ET measurements based on the area-weighted method, integrated Priestley-Taylor equation method and area-to-area regression kriging method are shown in Fig. 11.

In Fig. 11, the upscaled ET results from the area-weighted method exhibit relatively large errors, while the integrated Priestley-Taylor equation method and area-to-area regression kriging method exhibited good accuracy due to the use of remotely sensed LST. The MAPE of the upscaled ET estimates using the area-weighted method for September 12 was 23.2%, while the MAPE was 14.8% using the area-to-area regression kriging method and only 8.7% using the integrated Priestley-Taylor equation method. Therefore, upscaled ET results with high accuracy can be acquired by introducing auxiliary variables that reflect the surface water and heat conditions (i.e., remotely sensed LSTs) via an appropriate upscaling method, such as the integrated Priestley-Taylor equation method or area-to-area regression kriging method. The above work will be helpful to acquire "ground-truth" ET data at the satellite pixel scale over heterogeneous land surfaces.

4.5. Acquisition of "ground-truth" ET data at the satellite pixel scale

Based on the above analysis, the area-weighted method can be used to upscale multi-site ET measurements over homogeneous land surfaces. However, over heterogeneous land surfaces, auxiliary variables that can reflect the surface water and heat conditions (e.g., LSTs retrieved from remote-sensing images) should be introduced, after which the multi-site ET measurements can be upscaled via the integrated Priestley-Taylor equation method or area-to-area regression kriging method. Consequently, the precision of the upscaled results can be greatly improved. In the following analysis, a combined method was adopted to acquire instantaneous and daily "ground-truth" ET data at the satellite pixel scale during the timing of a MODIS overpass (applied the area-weighted method for relative homogeneous surface, otherwise used the method introducing auxiliary variables). The "ground-truth" ET data at the satellite pixel scale can be used to validate remotely sensed ET products and improve ET estimation models.

The 30-min averaged latent heat fluxes from multi-site EC measurements were upscaled using the area-weighted averaging method to coincide with the timing of a MODIS overpass, and the

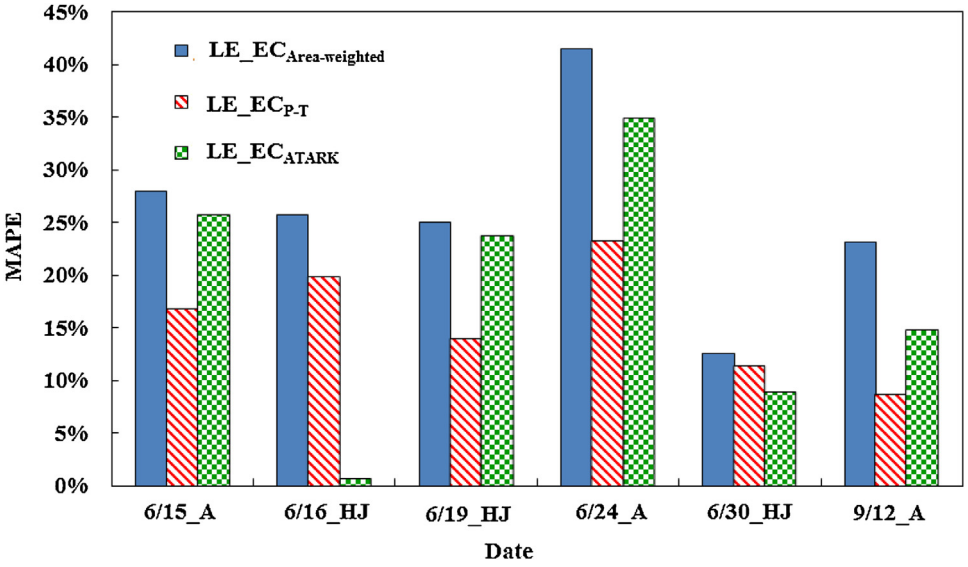

Fig. 11. Comparison of the upscaled LAS3 pixel results produced using the area-weighted method ($LE\_EC_{Area\text{-}weighted}$), the integrated Priestley-Taylor equation method ($LE\_EC_{P\text{-}T}$) and the area-to-area regression kriging method (LE_EC_{ATARK}; A: ASTER image; HJ: HJ-1B image) under conditions with high surface heterogeneity (before the crop was at full cover and at the end of the crop growing period).

instantaneous "ground-truth" ET data at the satellite pixel scale was obtained in the LAS1, LAS2, LAS3 and LAS4 pixels. Then, the MAPE thresholds of the upscaled ET data were set according to the current instantaneous estimation precision of the remotely sensed ET product (the MAPE was approximately 15–30%), the heterogeneity of the underlying surfaces and the remaining sample numbers. The MAPE threshold in the LAS1, LAS2, and LAS4 pixels was set to 6%, and the MAPE in the LAS3 pixels was set to 10%. Finally, if the MAPE of the upscaled ET data was larger than the threshold value, the integrated Priestley-Taylor equation method was adopted or set to the default value. The instantaneous "ground-truth" ET data in the LAS1, LAS2, LAS3 and LAS4 pixels were obtained.

As shown in Fig. 12a, the uncertainties (MAPE) in the instantaneous "ground-truth" ET data at the time of the MODIS overpass in the LAS1, LAS2, LAS3 and LAS4 pixels were 3.8%, 3.2%, 5.1%, and 3.1%, respectively. The upscaling precision of the latent heat flux was extremely high in the LAS1, LAS2, and LAS4 pixels, and was relatively poor in the LAS3 pixels, compared to the LAS observations. The uncertainties were primarily derived from the EC and LAS measurements and the upscaling process.

With regard to the daily "ground-truth" ET data at the satellite pixel scale at the time of the MODIS overpass, the area-weighted method was first used to upscale the multi-site ET measurements. Then, the MAPE thresholds of the upscaled ET data were set according to the current daily estimation precision of remotely sensed ET products (i.e., the MAPE was approximately 14–25%), the heterogeneity of underlying surfaces, and the remaining sample numbers. The MAPE threshold was 10% for the LAS1, LAS2, and LAS4 pixels and 13% for the LAS3 pixels. If the MAPE of the upscaled ET results was larger than the threshold value, the data were set to a default value. Finally, the daily "ground-truth" ET data in the LAS1, LAS2, LAS3 and LAS4 pixels were obtained. As shown in Fig. 12b, the uncertainty of daily "ground-truth" ET data at the time of the MODIS overpass in the LAS1, LAS2, LAS3 and LAS4 pixels was 4.7%, 5.4%, 10.7%, and 7.3%, respectively, compared to the LAS observations. The upscaled results were good in the LAS1, LAS2, and LAS4 pixels and relatively poor in the LAS3 pixels due to high levels of heterogeneity in the underlying surfaces. The uncertainties of daily "ground-truth" ET data at the time of the MODIS overpass primarily originated in the EC and LAS measurements and the upscaling process.

From the above analysis, the precision of the instantaneous and daily "ground-truth" ET data at the satellite pixel scale at the time of the MODIS overpass was good. The "ground-truth" ET data can therefore meet the requirements for validating remotely sensed ET products.

5. Conclusion

Based on multi-site measurements from 11 EC systems, 4 groups of LASs, and 3 groups of TDPs in three 3×1 (LAS1-3) MODIS pixels and one 2×1 MODIS pixels (LAS4) during the HiWATER-MUSOEXE, five upscaling methods, including three simple upscaling methods and two upscaling methods introducing auxiliary variables, were compared and a combined method was developed to acquire "ground-truth" ET data at the satellite pixel scale.

The three simple upscaling methods, namely, the arithmetic average method, area-weighted method, and footprint-weighted method, were compared by combining AWS data, remote-sensing data, and a footprint model. The results show that the simple upscaling methods can be applied to homogeneous underlying surfaces with relatively high precision, and the MAPE values in the LAS1, LAS2, and LAS4 pixels using the area-weighted averaging method were 7.7%, 5.4%, and 5.3% (an average of 6.1%), respectively. However, in the heterogeneous LAS3 pixels, the MAPE was 10.8%. Before full cover (before July 10) and at the end of the crop growing period (after September 4), the heterogeneity of the surface water and heat conditions was large, and the MAPE values of the LAS3 pixels for the two periods using the area-weighted averaging method were 16.3% and 10.0%, respectively. The MAPE of the LAS3 pixels dropped to 6.6% in the middle of the crop growing period (from July 10 to September 4), when the underlying surface was relatively homogenous. Therefore, the three simple upscaling methods cannot achieve satisfactory upscaled ET results over heterogeneous land surfaces. Consequently, auxiliary variables characterizing the surface heterogeneity should be introduced, and an appropriate upscaling method, such as the integrated Priestley-Taylor equation method or the area-to-area regression kriging method, can be used to improve the upscaled ET results.

On this basis, a combined method (applied the area-weighted method for relative homogeneous surface, otherwise used the method introducing auxiliary variables) was used to acquire instan-

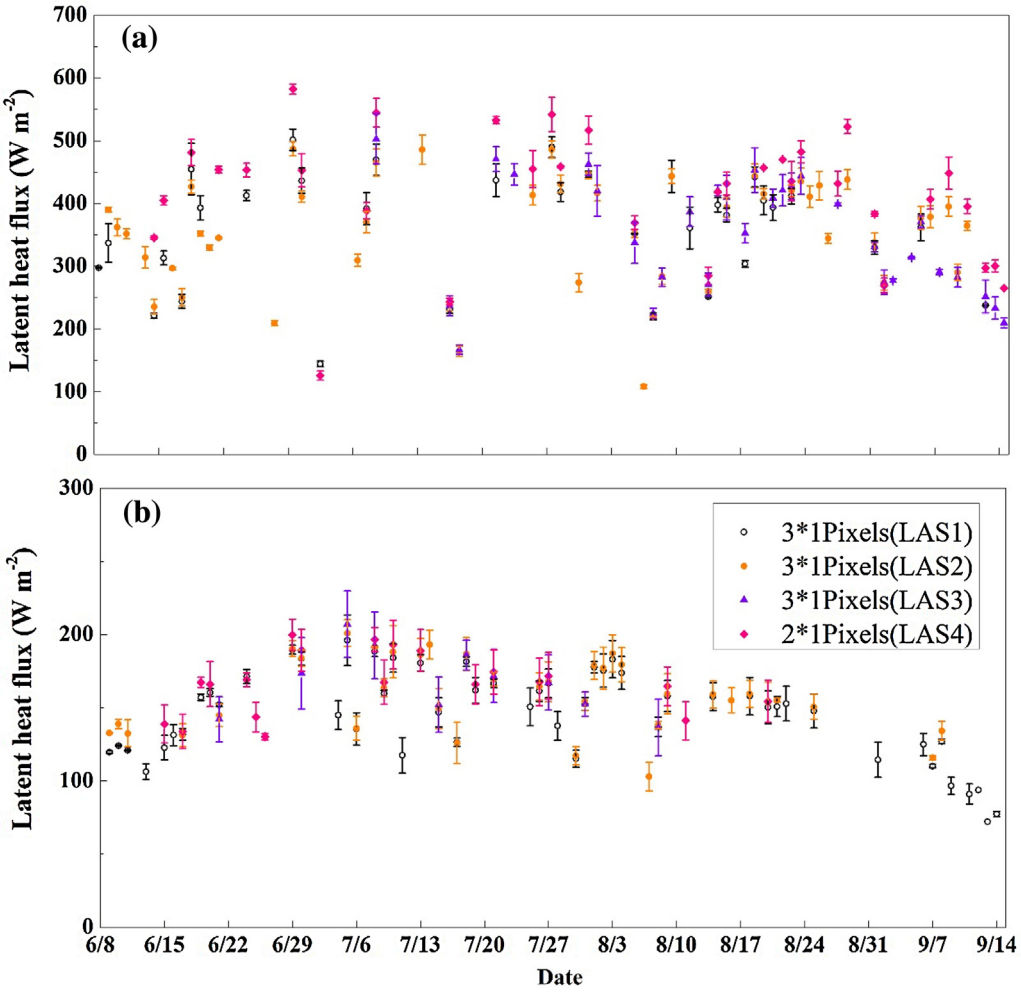

Fig. 12. Instantaneous and daily "ground-truth" ET data at the satellite pixel scale at the time of the MODIS overpass: (a) instantaneous and (b) daily. The error bar indicates the uncertainty in the "ground-truth" ET at the satellite pixel scale.

taneous and daily "ground-truth" ET data at the satellite pixel scale at the time of a MODIS overpass, and the uncertainty of the "ground-truth" ET data was evaluated, taking LAS measurements as the satellite pixel reference. Compared to the results of the corresponding observed LAS values, the uncertainties (MAPE) of the instantaneous "ground-truth" ET data for the LAS1, LAS2, LAS3 and LAS4 pixels were 3.8%, 3.2%, 5.1%, and 3.1%, respectively, and uncertainties (MAPE) of the daily "ground-truth" ET data were 4.7%, 5.4%, 10.7%, and 7.3%, respectively. The uncertainties of both the instantaneous and daily "ground-truth" ET data at the time of the MODIS overpass in the LAS1, LAS2, LAS3 and LAS4 pixels were primarily derived from the EC and LAS measurements and the upscaling process. The combined method to acquire "ground-truth" ET data at the satellite pixel scale could be used as a tool for bridging the gap between in situ ET measurements and remote-sensing estimates of ET.

In this study, problems were encountered due to the use of the LAS measurements as the "reference values" at the satellite pixel scale to evaluate the upscaled ET results from the multi-site EC measurements. The study area was in a desert-oasis system, which was affected by strong advection and secondary circulation. These conditions might result in small or negative sensible heat flux within the LAS1-4 pixels. An optical LAS can only measure the sensible heat flux, therefore, the latent heat flux was derived from the residual term of the surface energy balance equation or the relationship between the LAS-*LE* and area-averaged R_n. These errors associated with the LAS-derived sensible heat flux and latent heat flux when using the residual method requires further analysis. In addition, the integrated Priestley-Taylor equation method and the area-to-area regression kriging method for upscaling multi-site ET measurements over heterogeneous land surfaces require auxiliary variables to characterize the surface water and heat conditions, such as remotely sensed LSTs. However, these methods are only suitable for observations collected at the time of a satellite overpass due to the limitations of remote-sensing data. In future studies, multi-source data that can characterize surface water and heat conditions should be introduced to acquire daily "ground-truth" ET data at the satellite pixel scale over heterogeneous land surfaces.

Acknowledgements

We thank all the scientists, engineers and students who participated in the HiWATER-MUSOEXE.This work was supported by the National Natural Science Foundation of China (91125002), the National Basic Research Program of China (2015CB953702) and the National Natural Science Foundation of China (41531174).

References

Allen, R.G., Pereira, L.S., Howell, T.A., Jensen, M.E., 2011. Evapotranspiration information reporting: i: Factors governing measurement accuracy. Agric. Water Manage. 98 (6), 899–920.

André, J.C., Goutorbe, J.P., Perrier, A., 1986. HAPEX-MOBLIHY: a hydrologic atmospheric experiment for the study of water budget and evaporation flux at the climatic scale. Bull. Am. Meteorol. Soc. 67 (2), 138–144.
André, J.C., Bougeault, P., Goutorbe, J.P., 1990. Regional estimates of heat and evaporation fluxes over non-homogeneous terrain. Examples from the HAPEX-MOBILHY programme. Bound.–Lay. Meteorol. 50 (1-4), 77–108.
Baret, F., Clever, J.P.W., Stevens, M.D., 1995. The robustness of canopy gap fraction estimates from the red and near infrared reflectance: a comparison of approaches. Remote Sens. Environ. 54 (2), 141–151.
Berbigier, P., Bonnefond, J.M., Mellmann, P., 2001. CO_2 and water vapour fluxes for 2 years above Euroflux forest site. Agric. For. Meteorol. 108 (3), 183–197.
Beyrich, F., Mengelkamp, H.T., 2006. Evaporation over a heterogeneous land surface: EVA_GRIPS and the LITFASS-2003 experiment-an overview. Bound.–Lay. Meteorol. 121 (1), 5–32.
Beyrich, F., Herzog, H.J., Neisser, J., 2002. The LITFASS project of DWD and the LITFASS-98 experiment: the project strategy and the experimental setup. Theor. Appl. Climatol. 73 (1–2), 3–18.
Beyrich, F., Leps, J.P., Mauder, M., Bange, J., Foken, T., Huneke, S., Lohse, H., Lüdi, A., Meijninger, W.M.L., Mironov, D., Weisensee, U., Zittel, P., 2006. Area-averaged surface fluxes over the LITFASS region based on eddy-covariance measurements. Bound.–Lay. Meteorol. 121 (1), 33–65.
Davies, J.A., Allen, C.D., 1973. Equilibrium, potential and actual evaporation from cropped surfaces in southern Ontario. J. Appl. Meteorol. 12 (4), 649–657.
Doran, J.C., Hubbe, J.M., Kirkham, R.R., Shaw, W.J., Whiteman, C.D., Barnes, F.J., Cooper, D., Porch, W., Coutler, R.L., Cook, D.R., Hart, R.L., Gao, W., Martin, T.J., Shannon, J.D., Crawford, T.L., Baldcchi, D.D., Dobosy, R.J., Meyers, T.P., Balick, L., Dugas, W.A., Hicks, R., Fritschen, L., Hipps, L., Swiatek, E., Kunkel, K.F., 1992. The Boardman regional flux experiment. Bull. Am. Meteorol. Soc. 73 (11), 1785–1795.
Ezzahar, J., Chehbouni, A., Hoedjes, J.C.B., Chehbouni, A., 2007. On the application of scintillometry over heterogeneous grids. J. Hydrol. 334 (3), 493–501.
Ezzahar, J., Chehbouni, A., Hoedjes, J., Ramier, D., Boulain, N., Boubkraoui, S., Cappelaere, B., Descroix, L., Mougenot, B., Timouk, F., 2009. Combining scintillometer measurements and an aggregation scheme to estimate area-averaged latent heat flux during the AMMA experiment. J. Hydrol. 375 (1), 217–226.
Fisher, J.B., Tu, K.P., Baldocchi, D.D., 2008. Global estimates of the land–atmosphere water flux based on monthly AVHRR and ISLSCP-II data, validated at 16 FLUXNET sites. Remote Sens. Environ. 112 (3), 901–919.
Ge, Y., Liang, Y.Z., Wang, J.H., Zhao, Q.Y., Liu, S.M., 2015. Upscaling sensible heat fluxes with area-to-area regression kriging. IEEE Geosci. Remote Sens. Lett. 12 (3), 656–660.
Gottschalk, L., Batchvarova, E., Gryning, S.E., Lindroth, A., Melas, D., Motovilov, Y., 1999. Scale aggregation—comparison of flux estimates from NOPEX. Agric. For. Meteorol. 98–99, 103–119.
Goutorbe, J.P., Lebel, T., Tinga, A., Bessemoulin, P., Brouwer, J., Dolman, A.J., Engman, E.T., Gash, J.H.C., Hoepffner, M., Kabat, P., Kerr, Y.H., Monteny, B., Prince, S., Said, F., Sellers, P., Wallace, J.S., 1994. HAPEX-Sahel: a large-scale study of land-atmosphere interactions in the semi-arid tropics. Ann. Geophys. 12 (1), 53–64.
Gryning, S.E., Halldin, S., Lindroth, A., 2002. Area averaging of land surface-atmosphere fluxes in NOPEX: challenges, results and perspectives. Boreal Environ. Res. 7 (4), 379–388.
Halldin, S., Gryning, S.E., Gottschalk, L., Jochum, A., Lundin, L.C., Van de Griend, A.A., 1999. Energy, water and carbon exchange in a boreal forest landscape—NOPEX experiences. Agric. For. Meteorol. 98-99, 5–29.
Heinemann, G., Kerschgens, M., 2005. Comparison of methods for area averaging surface energy fluxes over heterogeneous land surfaces using high resolution non-hydrostatic simulations. Int. J. Climatol. 25 (3), 379–403.
Hoedjes, J.C.B., Chehbouni, A., Ezzahar, J., Escadafal, R., De Bruin, H.A.R., 2007. Comparison of large aperture scintillometer and eddy covariance measurements: can thermal infrared data be used to capture footprint-induced differences. J. Hydrometeorol. 8 (2), 144–159.
Jia, Z.Z., Liu, S.M., Xu, Z.W., Chen, Y.J., Zhu, M.J., 2012. Validation of remotely sensed evapotranspiration over the Hai River Basin, China. J. Geophys. Res. Atmos. 117, D13113, http://dx.doi.org/10.1029/2011jd017037.
Jiang, L., Islam, S., 2001. Estimation of surface evaporation map over southern Great Plains using remote sensing data. Water Resour. Res. 37 (2), 329–340.
Kalma, J., McVicar, T., McCabe, M., 2008. Estimating land surface evaporation: a review of methods using remotely sensed surface temperature data. Surv. Geophys. 29 (4–5), 421–469.
Kormann, R., Meixner, F.X., 2001. An analytical footprint model for non-neutral stratification. Bound.–Lay. Meteorol. 99 (2), 207–224.
Li, X., Li, X.W., Li, Z.Y., Ma, M.G., Wang, J., Xiao, Q., Liu, Q., Che, T., Chen, E.X., Yan, G.J., Hu, Z.Y., Zhang, L.X., Chu, R.Z., Su, P.X., Liu, Q.H., Liu, S.M., Wang, J.D., Niu, Z., Chen, Y., Jin, R., Wang, W.Z., Ran, Y.H., Xin, X.Z., Ren, H.Z., 2009. Watershed allied telemetry experimental research. J. Geophys. Res. Atmos. 114, D22103, http://dx.doi.org/10.1029/2008jd011590.
Li, X., Cheng, G.D., Liu, S.M., Xiao, Q., Ma, M.G., Jin, R., Che, T., Liu, Q.H., Wang, W.Z., Qi, Y., Wen, J.G., Li, H.Y., Zhu, G.F., Guo, J.W., Ran, Y.H., Wang, S.G., Zhu, Z.L., Zhou, J., Hu, X.L., Xu, Z.W., 2013. Heihe watershed allied telemetry experimental research (HiWATER): scientific objectives and experimental design. Bull. Am. Meteorol. Soc. 94 (8), 1145–1160.
Liebethal, C., Huwe, B., Foken, T., 2005. Sensitivity analysis for two ground heat flux calculation approaches. Agric. For. Meteorol. 132 (3), 253–262.
Liu, S.M., Xu, Z.W., Wang, W.Z., Jia, Z.Z., Zhu, M.J., Wang, J.M., 2011. A comparison of eddy-covariance and large aperture scintillometer measurements with respect to the energy balance closure problem. Hydrol. Earth Syst. Sci. 15 (4), 1291–1306.
Liu, S.M., Xu, Z.W., Zhu, Z.L., Jia, Z.Z., Zhu, M.J., 2013. Measurements of evapotranspiration from eddy-covariance systems and large aperture scintillometers in the Hai River Basin, China. J. Hydrol. 487, 24–38.
Lu, L., Liu, S.M., Xu, Z.W., Bai, J., Wang, J.M., 2010. The scale relationship of sensible heat flux measured by large aperture scintillometer and eddy covariance system. Adv. Earth Sci. (in Chinese) 25 (11), 1273–1282.
Meijninger, W.M.L., Hartogensis, O.K., Kohsiek, W., Hoedjes, J.C.B., Zuurbier, R.M., De Bruin, H.A.R., 2002. Determination of area-averaged sensible heat fluxes with a large aperture scintillometer over a heterogeneous surface–Flevoland field experiment. Bound.–Lay. Meteorol. 105 (1), 37–62.
Mu, X., Hu, R., Huang, S., Chen, Y., 2012. HiWATER: dataset of emissivity in the middle reaches of the heihe river basin in 2012. beijing normal university; cold and arid regions environmental and engineering research institute. Chin. Acad. Sci., http://dx.doi.org/10.3972/hiwater.042.2013.db.
Ochs, G.R., Wilson, J.J., 1993. A Second-generation Large Aperture Scintillometer, NOAA Tech. Memor. ERL ETL-232, NOAA Environmental Research Laboratories, Boulder, CO USA, pp. 24.
Peng, G.L., Liu, S.M., Cai, X.H., Lu, L., Xu, Z.W., 2008. Footprint analysis of turbulent flux measurement over heterogeneous surface. Chin. J. Atmos. Sci. (in Chinese) 32 (5), 1064–1070.
Priestley, C.H.B., Taylor, R.J., 1972. On the assessment of surface heat flux and evaporation using large-scale parameters. Mon. Weather Rev. 100 (2), 81–92.
Qiao, C., Sun, R., Xu, Z.W., Zhang, L., Liu, L.Y., Hao, L.Y., Jiang, G.Q., 2015. A study of shelterbelt transpiration and cropland evapotranspiration in an irrigated area in the middle reaches of the Heihe River in northwestern China. IEEE Geosci. Remote Sens. Lett. 12 (2), 369–373.
Sellers, P.J., Hall, F.G., Asrar, G., Strebel, D.E., Murphy, R.E., 1988. The first ISLSCP field experiment (FIFE). Bull. Am. Meteorol. Soc. 69 (1), 22–27.
Shuttleworth, W.J., Gurney, R.J., Hsu, A.Y., Ormsby, J.P., 1989. FIFE: the variation in energy partition at surface flux sites: remote sensing and large-scale global processes. IAHS Publ. 186, 67–74.
Shuttleworth, W.J., 1991. Insight from large-scale observational studies of land/atmosphere interactions. Surv. Geophys. 12, 3–30.
Timmermans, W.J., Su, Z., Olioso, A., 2009. Footprint issues in scintillometry over heterogeneous landscapes. Hydrol. Earth Syst. Sci. 13 (11), 2179–2190.
Twine, T.E., Kustas, W.P., Norman, J.M., Cook, D.R., Houser, P.R., Meyers, T.P., Prueger, J.H., Starks, P.J., Wesely, M.L., 2000. Correcting eddy-covariance flux underestimates over a grassland. Agric. For. Meteorol. 103 (3), 279–300.
Velpuri, N.M., Senay, G.B., Singh, R.K., Bohms, S., Verdin, J.P., 2013. A comprehensive evaluation of two MODIS evapotranspiration products over the conterminous United States: using point and gridded FLUXNET and water balance ET. Remote Sens. Environ. 139, 35–49.
Wang, K.C., Dickinson, R.E., 2012. A review of global terrestrial evapotranspiration: observation, modeling, climatology, and climatic variability. Rev. Geophys. 50, RG2005, http://dx.doi.org/10.1029/2011RG000373.
Wang, K.C., Liang, S.L., 2008. An improved method for estimating global evapotranspiration based on satellite estimation of surface net radiation, vegetation index, temperature, and soil moisture. J. Hydrometeorol. 9 (4), 712–727.
Xu, Z.W., Liu, S.M., Li, X., Shi, S.J., Wang, J.M., Zhu, Z.L., Xu, T.R., Wang, W.Z., Ma, M.G., 2013. Intercomparison of surface energy flux measurement systems used during the HiWATER-MUSOEXE. J. Geophys. Res. Atmos. 118 (23), 13140–13157.
Yao, Y.J., Liang, S.L., Cheng, J., Liu, S.M., Fisher, J.B., Zhang, X.D., Jia, K., Zhao, X., Qin, Q.M., Zhao, B., Han, S.J., Zhou, G.S., Zhou, G.Y., Li, Y.L., Zhao, S.H., 2013. MODIS-driven estimation of terrestrial latent heat flux in China based on a modified Priestley-Taylor algorithm. Agric. For. Meteorol. 171–172, 187–202.
Zhang, M., Jiang, Z.R., Ma, M.G., Wang, Z.H., Zhang, T.F., 2013. Fine classification of planting structure in the middle reaches of heihe river basin based on hyperspectral compact airborne spectrographic imager (CASI) data. Remote Sens. Technol. Appl. (in Chinese) 28 (2), 283–289.
Zhou, J., Li, M.S., Liu, S.M., Jia, Z.Z., Ma, Y.F., 2015. Validation and performance evaluations of methods for estimating land surface temperatures from ASTER data in the middle reach of the Heihe River Basin. Northwest China Remote Sens. 7, 7126–7156.